Robot 2015, Second Iberian Robotics Conference, Vol.418 pp.259-271

Influence of Positive Instances on Multiple Instance Support Vector Machines

Nuno Barroso Monteiro[1,2], João Pedro Barreto[2], and José Gaspar[1]

[1] Institute for Systems and Robotics (ISR/IST), LARSyS, Univ. of Lisbon, Portugal
[2] Institute for Systems and Robotics, Univ. of Coimbra, Portugal
nmonteiro@isr.ist.utl.pt, jpbar@isr.uc.pt, jag@isr.ist.utl.pt

Abstract. This work studies the influence of the percentage of positive instances on positive bags on the performance of multiple instance learning algorithms using support vector machines. There are several studies that compare the performance of different types of multiple instance learning algorithms in different datasets and the performance of these algorithms with the supervised learning counterparts. Nonetheless, none of them study the influence of having a low or high percentage of positive instances on the data that the classifiers are using to learn. Therefore, we have created a new image dataset with different percentages of positive instances from a dataset for pedestrian detection. Experimental results of the performance of mi-SVM and MI-SVM algorithms on an image annotation task are presented. The results show that higher percentages of positive instances increase the overall accuracy of classifiers based on the maximum bag margin formulation.

Keywords: Multiple Instance Learning, mi-SVM, MI-SVM

1 Introduction

In supervised learning, the classifier is provided with a training set that consists of instances and the corresponding labels. The training set is then used to obtain a classifier that can predict the labels for novel instances [4]. Nonetheless, the correspondence requirement between instances and labels is difficult or even prohibitive for some applications like object detection. Annotating whole images is easier and faster than annotating and identifying relevant image regions.

Multiple instance learning appeared as a more flexible paradigm assuming that there is some ambiguity in how the labels are assigned. Namely, in multiple instance learning, the instances are grouped into bags and the labels are assigned to the bags instead of being assigned to each of the instances. The labels are then learned using a multiple instance learning assumption like the weighted collective assumption [10] or the standard multiple instance learning assumption. The standard multiple instance learning assumption is the one that is considered in this work, and states, for a binary classification problem: a bag is positive if at least one of the instances in that bag is positive, a bag is negative if all the instances in that bag are negative. Therefore, the true input labels are not known during training, i.e., the true input labels are latent variables. Note also that a positive bag can contain negative instances.

In order to understand better the multiple instance learning problem, let us consider an example adapted from [7]. Consider that we have a classroom and we know some professors that have access to that classroom, and others who do not have access. Each professor has a key chain with a few keys that can be differentiated, for example, by color. The goal is to predict if a given key or a given key chain allows to open the door of the classroom. Using the multiple instance learning framework, the bags will correspond to the key chains that are labeled as positive or negative according to the access that the professor has to the classroom. The instances are the keys contained in the key chains. Using the assumption, we know that keys on key chains from professors that do not have access to the classroom do not open the door. Thus, to solve this problem we have to find the key that is common in all positive key chains. If we consider the color as a feature, the keys colors that appear in negative key chains could be ruled out. Hopefully, there is one key color that remains and this will correspond to the key that opens the door to the classroom. If the classifier can correctly identify this key, it can predict if a key or key chain is able to give access to the classroom. From this simple example, we can see that multiple instance learning algorithms can have several formulations because they can aim at designing classifiers for better discriminating bags or instances.

The performance of multiple instance learning algorithms have been compared between themselves and with the supervised learning counterparts in several application domains. But none of them focused on what is the influence of a different percentage of positive instances on the classifiers. Therefore, this work intends to study this influence on the performance of the classifiers. We focus on mi-SVM and MI-SVM which are support vector machines adapted to the multiple instance learning framework.

The paper has two major contributions: a new dataset that allows to evaluate the influence of different percentages of positive instances, and experimental evidence that increasing this percentage has a positive impact in the performance of MI-SVM.

In terms of structure, we will first present a brief review of the state of the art on multiple instance learning algorithms in Section 2. A more detailed formalism of multiple instance learning frameworks with focus on algorithms using support vector machines is provided in Section 3. The methodology used to evaluate the performance of the classifiers and the dataset used is described in Section 4. The results and major conclusions are presented in Section 5 and Section 6, respectively.

Notation: The notation followed throughout this work is the following: non-italic letters correspond to functions, italic letters correspond to scalars, lower case bold letters correspond to vectors (for example, to represent instances), and upper case bold letters correspond to matrices (for example, to represent bags).

2 Related Work

Multiple instance learning algorithms appeared to overcome some limitations of supervised learning regarding the training sets for some applications like object detection or drug activity prediction. The limitation is associated with the difficulty of providing an accurate and correctly labeled training set at the instance level.

The term multiple instance learning appeared with Dietterich *et al.* [7] in the context of drug activity prediction. Dietterich *et al.* [7] developed the Axis-Parallel Rectangle algorithm for predicting some property of the molecule based on the molecule's shape statistics. Since than, several algorithms have been proposed for solving the multiple instance learning problem. Maron *et al.* [13] proposed the Diverse Density (DD) algorithm to learn Gaussian concepts for representing the positive regions. This algorithm was then extended by Zhang *et al.* [17] to use the expectation-maximization (EM) has the optimization technique.

Other authors have proposed algorithms for multiple instance learning by adapting the support vector machine framework. The support vector machine framework aims in finding an hyperplane that is capable of separating the training data with the maximum margin possible. Andrews *et al.* [2] proposed two algorithms, mi-SVM and MI-SVM, that differ in the margin definition. The first considers the margin between the positive and negative instances while the second considers the margin between the bags. In the MI-SVM, the bags are not represented by multiple instances but rather by the most positive instance for the positive bags and by the least negative instance for the negative bags. Bunescu *et al.* [5] extended the MI-SVM algorithm by adding constraints to ensure that at least one of the instances in a positive bag is positive. This algorithm is shown to work well for sparse positive bags.

The previous approaches modify the objective function for adapting the support vector machines to the multiple instance learning framework. There are other approaches that modify the kernels used. Gartner *et al.* [11] proposed two kernels: statistic kernel and the normalized set kernel. In the statistic kernel, the bag is transformed into a feature vector by selecting the minimum and maximum values for each feature from all instances in the bag. In the normalized set kernel, the bag is represented as the sum of all instances that belong to the bag normalized by the 1 or 2-norm.

These algorithms have been applied in different application domains like classification of molecules [7], content based image retrieval [2], text classification [2], among others. These algorithms have different formalisms and a more detailed review can be found in the following articles [1, 3, 10].

Since there are a large number of applications and a high number of algorithms for multiple instance learning, several studies have concentrated in finding the best multiple instance classifier. Ray and Craven [16] concluded that there is no optimal multiple instance learning algorithm since their performance depend on the data. Furthermore, the multiple instance learning algorithms have been compared to the correspondent supervised learning algorithms using different datasets [5, 7, 16]. These studies concluded that the multiple instance learning algorithms are consistently superior, although this observation also depends on the data. Nonetheless, none of these studies have considered the rate of positive instances in their data while evaluating the performance of the multiple instance learning algorithms. The rate of positive instances on positive bags is directly related with the standard multiple instance learning assumption. Therefore, in this work we want to evaluate if the rate of positive instances on positive bags influence the performance of multiple instance learning algorithms. This will allow to determine if this should be considered as a variable when analyzing the performance of these algorithms.

3 Multiple Instance Learning

In this section we will present the formalism associated with the multiple instance learning problem. Remember from Section 1 that in supervised learning, the training set consists of example pairs: an input and a corresponding label. In multiple instance learning, the inputs are grouped into bags and the labels are assigned to the bags of inputs. Thus, we do not know which of the inputs or pair of inputs is responsible for the label. In this sense, the multiple instance learning framework makes weaker assumptions about the labeling information.

In order to formulate the multiple instance problem, let us consider a training data $\mathbf{\Phi}$ with N pairs of examples:

$$\mathbf{\Phi} = \{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \ldots, (\mathbf{X}_N, y_N)\} \tag{1}$$

where $\mathbf{X}_i$ is the bag of the i-th example, and y_i is the corresponding bag label of the i-th example. A bag $\mathbf{X}_i$ is composed of M_i instances:

$$\mathbf{X}_i = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iM_i}\} \tag{2}$$

where $\mathbf{x}_{ij} \in \chi_j$ is the j-th instance of the bag $\mathbf{X}_i$ and $\chi_j = \mathbb{R}^{D_j}$ is the D_j-dimensional Euclidean space (dimension of the j-th instance). For simplifying the notation, assume from now on that all bags in the N pairs of examples have the same number of instances $M_i = M$ and all instances belong to the same D-dimensional Euclidean space $\chi_j = \chi = \mathbb{R}^D$.

A bag label $y_i \in \Upsilon$ is the result of the labels given to each of the instances $y_{ij} \in \upsilon$ that compose the bag $\mathbf{X}_i$. For simplifying this exposure, consider a binary classification problem where $\Upsilon = \{-1, 1\}$ and $\upsilon = \{-1, 1\}$. Using the standard multiple instance learning assumption for binary classification, mentioned in Section 1, the bag label y_i corresponds to:

$$y_i = \max_j y_{ij} = \begin{cases} 1 & \exists_j : y_{ij} = 1 \\ -1 & \forall j : y_{ij} = -1 \end{cases} \tag{3}$$

The goal of the multiple instance learning algorithm is to train an instance classifier $\mathrm{f}(\mathbf{x}) : \chi \to \upsilon$ or a bag classifier $\mathrm{F}(\mathbf{X}) : \chi^M \to \Upsilon$. From (3), a bag classifier can be obtained from a correct instance classifier by $\mathrm{F}(\mathbf{X}_i) = \mathrm{sign}(\mathrm{max_j}\, \mathrm{f}(\mathbf{x}_{ij}))$. Hence, most of the multiple instance learning algorithms aim to learn instance classifiers instead of bag classifiers.

The focus of this work is to evaluate the performance of support vector machines in the multiple instance learning framework in an image classification task, therefore we will now present the formalism of the two classifiers used: Maximum Instance Margin (mi-SVM) and Maximum Bag Margin (MI-SVM). But first, let us introduce the support vector machine framework for supervised learning.

3.1 Support Vector Machines

Consider a training set $\mathbf{\Omega}$ with K examples:

$$\mathbf{\Omega} = \{(\mathbf{x}_{i1}, y_{i1}), (\mathbf{x}_{i2}, y_{i2}), \ldots, (\mathbf{x}_{iK}, y_{iK})\} \tag{4}$$

in a binary classification problem $y_{ij} = \{-1, 1\}$. The objective of a support vector machine framework is to find the hyperplane that separates the examples with the biggest margin possible. The margin is defined as the smallest distance between the hyperplane and a positive and a negative example.

The support vector machine soft-margin formulation is:

$$\begin{aligned} &\min_{\mathbf{w}, w_0, \xi_i} \tfrac{1}{2} \|\mathbf{w}\|^2 + C \sum_{j=1}^{K} \xi_{ij} \\ &s.t. \quad \xi_{ij} \geq 0 \\ &\qquad\; y_{ij} \left(\mathbf{w} \cdot \mathbf{x}_{ij} + w_0 \right) \geq 1 - \xi_{ij} \end{aligned} \tag{5}$$

where $\mathbf{w}$ is the hyperplane normal, w_0 is the hyperplane offset, and the ξ_{ij} are the slack variables that allow to apply the support vector machine framework to data that is not separable. This means that we are allowed some mislabeled examples. This leads to a quadratic programming problem that is convex and easily solvable. The examples that are nearest to the hyperplane are called the support vectors.

In this work we do not intend to detail the support vector machine formalism. Therefore, for a detailed introduction to support vector machines the reader should refer to [12, 14].

3.2 Maximum Instance Margin: mi-SVM

The maximum instance margin formulation of the support vector machines aims to recover the instance labels of the positive bags.

In support vector machines for supervised learning, the labels y_{ij} of each instance $\mathbf{x}_{ij}$ in a training set are known. In multiple instance learning this is not the case, only the labels y_i of a bag $\mathbf{X}_i$ are known. Considering the standard multiple instance learning assumption (3), we can see that the labels for each instance of a negative bag are also known and therefore the margin could be defined as in a regular support vector machine. However, the labels for each instance of a positive bag are unknown and therefore computing the margin is more complicated. Andrews *et al.* [2] treated the instance labels y_{ij} as unknown integer variables:

$$\begin{aligned} &\min_{y_{ij}} \min_{\mathbf{w}, w_0, \xi_i} \tfrac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i,j} \xi_{ij} \\ &s.t. \quad \xi_{ij} \geq 0 \\ &\qquad\; y_{ij} \left(\mathbf{w} \cdot \mathbf{x}_{ij} + w_0 \right) \geq 1 - \xi_{ij} \\ &\qquad\; y_{ij} = -1, \ \forall i : y_i = -1 \\ &\qquad\; \sum_{j} \tfrac{y_{ij}+1}{2} \geq 1, \ \forall i : y_i = 1 \end{aligned} \tag{6}$$

From (6) we can see that in mi-SVM multiple negative or positive instances in positive bags can be support vectors. The problem leads to a mixed integer program that is hard to solve. Nonetheless, the integer variables, the hidden labels y_{ij}, reduce the problem to a quadratic programming problem. Andrews *et al.* [2] proposed an heuristic that consists of two steps: first train a SVM classifier considering a given value for the instance labels, then use the new classifier to update the instance labels y_{ij} of positive bags. This process is computed until no changes occur in the instance labels.

3.3 Maximum Bag Margin: MI-SVM

The maximum bag margin formulation of the support vector machines aims to recover the key positive instance for every positive bag.

In this formulation, the margin of a bag corresponds to the maximum distance between the hyperplane and all of the instances that belong to a bag:

$$y_i = \max_j \left(\mathbf{w} \cdot \mathbf{x}_{ij} + w_0\right) \tag{7}$$

From (7) we can see that the margin of a positive bag is determined by the most positive instance while the margin of a negative bag is determined by the least negative instance. The bag margin formulation is:

$$\begin{aligned} &\min_{\mathbf{w}, w_0, \xi_i} \tfrac{1}{2} \left\|\mathbf{w}\right\|^2 + C \sum_{i,j} \xi_{ij} \\ &s.t. \quad \xi_{ij} \geq 0 \\ &\qquad y_i \max_j \left(\mathbf{w} \cdot \mathbf{x}_{ij} + w_0\right) \geq 1 - \xi_{ij} \end{aligned} \tag{8}$$

This optimization is not convex, therefore Andrews *et al.* [2] introduced an extra variable $s(j)$ for each bag. This variable denotes the instance that is selected as the witness instance of a positive bag. The formulation (8) is now given by:

$$\begin{aligned} &\min_{s(j)} \min_{\mathbf{w}, w_0, \xi_i} \tfrac{1}{2} \left\|\mathbf{w}\right\|^2 + C \sum_{i,j} \xi_{ij} \\ &s.t. \quad \xi_{ij} \geq 0 \\ &\qquad \mathbf{w} \cdot \mathbf{x}_{ij} + w_0 \leq -1 + \xi_{ij}, \forall i : y_i = -1 \\ &\qquad \mathbf{w} \cdot \mathbf{x}_{is(j)} + w_0 \geq 1 - \xi_{ij}, \forall i : y_i = 1 \end{aligned} \tag{9}$$

In this formulation, each positive bag is represented by only one positive instance. All the negative instances in the positive bags are disregarded. Like the formulation of mi-SVM, this corresponds to a mixed integer program. Nonetheless, the integer variables, the variables $s(j)$, reduce the problem to a quadratic programming problem. Andrews *et al.* [2] also proposed an heuristic that consists in two steps: first train a classifier like a regular supervised learning considering a given witness instance, then using the new classifier select new witness instances for the positive bags. The optimization process ends when the selected witness stops to change.

4 Methodology

The objective of this work is to compare the performance of multiple instance support vector machines' learning algorithms with the percentage of positive instances on positive bags. This performance is evaluated in an image annotation task using a dataset for pedestrian detection. Image annotation task consists on identifying if an image has a person and if a given region of that image contains a person or a part of a person.

Normally, datasets in multiple instance learning do not report the percentage of positive instances included on the positive bags and there is no study evaluating the

Fig. 1. Examples of positive and negative instances obtained during the transformation of the HDA Person Dataset. A and C: Negative instances (cyan) drawn from negative bags. B and D: Negative (cyan) and positive (red) instances drawn from positive bags.

influence of positive instances on this type of classifiers. Therefore, a new multiple instance learning dataset has been created based on a real dataset for pedestrian detection, the HDA Person Dataset [9, 15].

The HDA dataset is a high resolution image sequence dataset for research on high definition surveillance, pedestrian detection and re-identification. The dataset comprises information from 18 cameras with different resolutions. A total of 13 image sequences are labeled. The labeled data includes 64.028 annotations from 85 persons in a total of 75.207 frames. The annotations include information about the person bounding box position and unique identification, occlusion and type of detection (person or crowd). Additionally, the annotations have information about the camera and frame number.

In order to transform this dataset, we adopted a similar strategy to Andrews *et al.* [2]. This takes in consideration the standard multiple instance learning assumption for binary classification: an image is positive if at least there is one person or a part of a person on the image, and an image is negative if there is no person or part of a person on the image. Therefore, the positive images are randomly drawn from the annotated frames of the cameras while the negative images are sampled from the non-annotated frames. Notice that frames annotated with occlusion are not considered, in order to not provide erroneous features to the classifier. Remember that the bounding box for an occluded person is drawn by estimating the whole body extent in the HDA Person Dataset.

In a positive image (Figure 1.B and Figure 1.D), the positive image regions are obtained by sampling randomly the bounding box area while the negative image regions are drawn from the remaining area of the annotated frames. In a negative image (Figure 1.A and Figure 1.C), the negative image regions are randomly sampled from the entire

Set	Positive Instance Rate	Positive Bags		Negative Bags	Total Bags	
		Total Instances	Positive Instances	Negative Instances	Total Instances	Positive Instances
Testing	0.469	409	192	500	909	192
Training-10	0.100	750	75	750	1500	75
Training-30	0.300	750	225	750	1500	225
Training-50	0.500	750	375	750	1500	375
Training-70	0.699	747	522	750	1497	522
Training-90	0.899	739	664	750	1489	664

Table 1. Number of instances for each training and testing set obtained from the HDA Person Dataset.

area of the non-annotated frame. In either of the cases, the maximum overlapping area between image region of the same image is 40%, and the window considered for each image region is 60 x 60. Furthermore, in order to comply with a given percentage of positive image regions on positive images, only annotated frames that satisfied this percentage have been considered for each of the training datasets obtained. Like the MIL datasets mentioned previously, the number of instances per bag is also variable. The features are extracted from the image regions using the integral channel features defined by Dollar *et al.* [8] and that have been used for pedestrian detection. These features include color (LUV color space), gradient magnitude, and gradient histograms. Remember from Section 3 that each instance belongs to a D-dimensional Euclidean space, a feature corresponds to each of the D components of that Euclidean space. The number of features per image region is 600.

The training and testing sets have been obtained using the approach defined above. A total of 5 training sets were created, each with a different percentage of positive instances on positive bags: 0.1, 0.3, 0.5, 0.7, and 0.9. In order to have an unbiased comparison, the testing set consists of sets with different percentage of positive instances: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. Each of these sets are composed of approximately the same number of bags. Notice also that the images for the training and testing sets were obtained from different cameras. The training set was obtained from cameras 53, 56, 57 and 58, while the testing set was obtained from cameras 50 and 59. The images used correspond to cameras of equal resolution 1280 x 800.

In conclusion, the HDA Person dataset was transformed into a new dataset with training sets that consist of 150 bags (75 positive and 75 negative bags), and into a testing set of 100 bags (50 positive and 50 negative bags). The summarized statistics of each dataset obtained can be found in Table 1.

Similarly to the previous datasets, this dataset is used for training and testing the mi-SVM and MI-SVM classifiers. The parameters for each of the kernels (linear, polynomial and RBF [6]) were analyzed and optimized using 3-fold cross-validation using a training set of 150 bags (75 positive and 75 negative bags) obtained similarly to the testing set described above.

5 Results

The classifier and kernels accuracies depend on the data and on the application domain [16]. Therefore, we started by evaluating the performance of each of the classifiers and kernels on a specific training set (Training-CL) before analyzing the influence of the positive instances on positive bags. This training set corresponds to a mix of bags with different percentages of positive instances in order to get training data that is not biased towards one of the training sets with a specific percentage of positive instances. This analysis will allow us to select the best classifiers and kernels to the problem of pedestrian detection. The results obtained are presented in Table 2.

From Table 2, we can conclude that the mi-SVM algorithms are consistently better than the MI-SVM counterparts. This means that the features present at the instance level are sufficient to learn a classifier that can discriminate between the positive and negative bags. Furthermore, we can conclude that the mi-SVM classifier with RBF kernel has the highest bag and instance classification accuracies for the adapted HDA Person dataset. Notice also that the instance accuracies for mi-SVM with a linear kernel are similar to the ones obtained using a RBF kernel. This analysis allows us to determine that the classifiers for this data should use RBF kernels. Therefore, the remaining analysis were made using the mi-SVM classifier with RBF kernel and the MI-SVM classifier with RBF kernel.

As mentioned in Section 4, the performance of the mi-SVM and MI-SVM classifiers is evaluated for different percentages of positive instances on positive bags. The classification accuracies obtained are reported in Table 3. From Table 3, the increasing percentage of positive instances has higher influence on the maximum bag margin algorithms than on the maximum instance margin algorithms. In the mi-SVM, the instance classification accuracies are very similar for the different percentages of positive instances (maximum difference between the minimum and maximum classification accuracies is 1.8%), and the bag classification accuracy reaches a maximum when the training set has 50% of positive instances and then starts to decrease. In the MI-SVM, the bag and the instance classification accuracies increase significantly with the increasing percentage of positive instances.

The results in Table 3 give information about the overall accuracy of the classifier. In Figure 2, we present the performance of the classifier on positive and negative instances. From Figure 2, we can reinforce that the impact of an increasing percentage of positive instances is higher on MI-SVM classifier. Namely, when the rate of positive instances increases the classification accuracy of positive bags and instances increases. Nonetheless, this increase is followed by a decrease in the classification accuracy of negative bags and instances. This decrease occurs at a significantly less extent on instances than on bags. Regarding the mi-SVM classifiers, the influence of positive instances is more noticeable at the bag level. The change in the classification accuracy is driven by the change on the classification accuracy of negative bags (maximum difference between the minimum and maximum classification accuracies is 11%). The results show that the classifier is biased towards negative bags when there is a low and high percentage of positive instances. At the instance level, the accuracy does not have significant changes with the increase of positive instances.

In conclusion, for maximum bag margin algorithms, the increase of positive instances allow to obtain classifiers with higher accuracy (Figure 3). Nonetheless, at the bag level, the overall accuracy of the classifier does not benefit when the percentage of positive instances is too high (90%). These findings can be explained by the fact that

Training Set	Bag Classification Accuracy						Instance Classification Accuracy					
	mi-SVM			MI-SVM			mi-SVM			MI-SVM		
	Linear	**Polynomial**	**RBF**	**Linear**	**Polynomial**	**RBF**	**Linear**	**Polynomial**	**RBF**	**Linear**	**Polynomial**	**RBF**
Training-CL	0.680	0.510	**0.770**	0.600	0.500	0.710	0.878	0.677	**0.879**	0.804	0.789	0.816

Table 2. Bag and instance classification accuracies on an adapted HDA Person dataset. The higher classification accuracies are presented in bold.

Training Sets	mi-SVM.RBF		MI-SVM.RBF	
	Bag Accuracy	**Instance Accuracy**	**Bag Accuracy**	**Instance Accuracy**
Training-10	0.730	0.882	0.500	0.789
Training-30	0.760	**0.889**	0.500	0.789
Training-50	**0.790**	0.871	0.580	0.795
Training-70	0.780	0.882	**0.760**	0.843
Training-90	0.750	0.879	0.750	**0.864**

Table 3. Bag and instance classification accuracies for the mi-SVM and MI-SVM with RBF kernel obtained from training in datasets with different rates of positive instances.

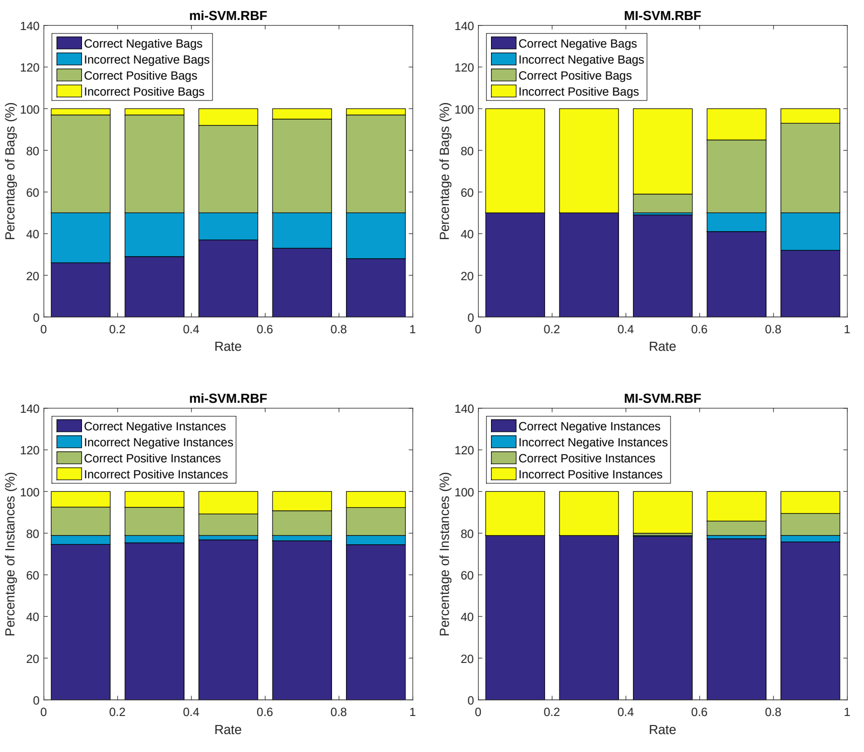

Fig. 2. Percentages of correctly and incorrectly classified bags (top) and instances (bottom) with increasing percentage of positive instances on positive bags for mi-SVM (left) and MI-SVM (right) classifiers.

these multiple instance algorithms use the most positive instance of each positive bag and the least negative of each negative bag to find the hyperplane that best separate the two types of instances. Thus, the higher percentage of positive instances allows the classifier to span more hypothesis for finding the instance that is most representative of the positive class. Nonetheless, after finding the witness instance of the positive class, adding more instances would not help increase the accuracy of the classifier. These results suggest that is better to use images whose random sampling has higher likelihood of providing positive instances.

Fig. 3. Example of an instance classification using the MI-SVM classifier for increasing rate of positive instances on positive bags: 0.5, 0.7, and 0.9. The image regions in green correspond to instances correctly labeled while the image regions in red correspond to mislabeled instances.

On the other hand, for maximum instance margin algorithms, the increase of positive instances do not exhibit any influence on the classifier accuracy at the instance level. This may be due to the fact that these algorithms use multiple positive or negative instances to determine the hyperplane. Although the percentage of positive instances is increasing, the negative instances are present in a high percentage in all training sets. Therefore, these classifiers are highly capable of determining the negative instances from the remaining instances. This could justify why the bag and instance accuracy of these classifiers is high throughout the several rate of positive instances. Nonetheless, at the bag level, the classifier shows signals of overfitting for high and low rate of positive instances.

6 Conclusions

In this work, the influence of positive instances on positive bags is analyzed on support vector machines adapted for the multiple instance learning framework: mi-SVM and MI-SVM. The algorithms were evaluated using a new dataset obtained from an existing pedestrian detection dataset.

The results show that the increasing percentage of positive instances have a positive impact in the overall accuracy of the MI-SVM classifier as a result of the maximum bag margin formulation. For the mi-SVM classifiers, no relevant changes occur at instance level. At bag level, the classifier is biased towards negative bags at low and high rates of positive instances affecting the overall performance of the classifier. Therefore, the rate of positive instances should be considered while evaluating the performance of mi-SVM and MI-SVM classifiers.

As future work, we would like to extend this analysis to more multiple instance learning algorithms and consider algorithms that are based on other multiple instance learning assumptions.

Acknowledgments

This work has been partially supported by the Portuguese Foundation for Science and Technology (FCT) project [UID / EEA / 50009 / 2013], by the CMU-Portugal Project AHA [CMUP-ERI / HCI / 0046 / 2013], and by the EU Project POETICON++ EU-FP7-ICT-288382. Nuno Barroso Monteiro is funded by FCT PhD grant PD/BD/105778/2014.

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. 561–568 (2002)
3. Babenko, B.: Multiple instance learning: algorithms and applications. View Article PubMed/NCBI Google Scholar (2008)
4. Bishop, C.M.: Pattern recognition and machine learning. springer (2006)
5. Bunescu, R.C., Mooney, R.J.: Multiple instance learning for sparse positive bags. In: Proceedings of the 24th international conference on Machine learning. pp. 105–112. ACM (2007)
6. Burges, C.J.: A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2(2), 121–167 (1998)
7. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artificial intelligence 89(1), 31–71 (1997)
8. Dollár, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: BMVC. vol. 2, p. 5 (2009)
9. Figueira, D., Taiana, M., Nambiar, A., Nascimento, J., Bernardino, A.: The hda+ data set for research on fully automated re-identification systems. In: Computer Vision-ECCV 2014 Workshops. pp. 241–255. Springer (2014)
10. Foulds, J., Frank, E.: A review of multi-instance learning assumptions. The Knowledge Engineering Review 25(01), 1–25 (2010)
11. Gärtner, T., Flach, P.A., Kowalczyk, A., Smola, A.J.: Multi-instance kernels. In: ICML. vol. 2, pp. 179–186 (2002)
12. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learnin (2009)
13. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. Advances in neural information processing systems pp. 570–576 (1998)
14. Murphy, K.P.: Machine learning: a probabilistic perspective. MIT press (2012)
15. Nambiar, A., Taiana, M., Figueira, D., Nascimento, J.C., Bernardino, A.: A multi-camera video dataset for research on high-definition surveillance. International Journal of Machine Intelligence and Sensory Signal Processing 1(3), 267–286 (2014)
16. Ray, S., Craven, M.: Supervised versus multiple instance learning: An empirical comparison. In: Proceedings of the 22nd international conference on Machine learning. pp. 697–704. ACM (2005)
17. Zhang, Q., Goldman, S.A.: Em-dd: An improved multiple-instance learning technique. In: Advances in neural information processing systems. pp. 1073–1080 (2001)