

**TITLE:** MIRSVM: Multi-Instance Support Vector Machine with Bag Representatives

**AUTHORS:** Gabriella Melki, Alberto Cano, and Sebastián Ventura

Machine learning contribution information:

- ***What is the main claim of the paper? Why is this an important contribution to the machine learning literature?***

This paper proposes a novel L1 support vector machine (SVM) formulation, with a bag-representative selector, called Multiple-Instance Representative Support Vector Machine (MIRSVM). Our approach was based on how SVMs identify key instances (support vectors) that highly impact classification using bag-level information. Combining SVMs with a bag-representative selector aims to remedy potential bias towards the negative class issues present in many multiple-instance learning approaches. The proposed method makes no assumptions regarding the distribution of the instances within positive bags caused by methods such as wrappers. This ensures all the information contained within each bag is properly utilized during training.

- ***What is the evidence you provide to support your claim?***

MIRSVM's bag-representative selector is designed to optimize over both positive and negative bags, ensuring each class is properly accounted for during learning. The primal L1-SVM problem was reformulated to optimize the margin over bag-representatives, using bag-level information. The dual was then derived and solved using a quadratic programming solver, producing a sparse model. MIRSVM's performance was evaluated with the usage of multiple metrics: accuracy, precision, recall, Cohen's Kappa rate, and Area Under Receiver Operating Characteristic curve (AUC) and the results indicate that the representative selection process contributes to the algorithm's better performance against the state-of-the-art algorithms. MIRSVM achieves the best average efficacy over the 15 benchmark datasets used, especially in comparison to current multi-instance SVM methods. MIRSVM outperforms the ensemble methods used in the experiments according to recall, despite them exhibiting relatively good accuracy and precision indicating them being strongly conservative towards predicting positive bags. MIRSVM's kappa values all fall within the range (0.5-1], highlighting its merit as a classifier. This is also supported by the AUC results which shows a high true positive rate, which is reflected by the precision and recall results. Even though MIRSVM uses a quadratic programming solver to compute the dual SVM problem, it still outperforms all the algorithms compared in terms of run-time, especially over datasets with large number of features and instances.

- ***What papers by other authors make the most closely related contributions, and how is your paper related to them?***

The most closely related contributions include:

- G. Doran and S. Ray. A theoretical and empirical analysis of support vector machine methods for multiple-instance classification. *Machine Learning*, 97(1-2):79-102, 2014.
- S. Andrews, I. Tsochantaridis, and T. Hofmann. Support Vector Machines for Multiple-Instance Learning. *15th International Conference on Neural Information Processing Systems*, 577-584, 2002.
- A.W.C. Faria, F.G.F. Coelho, A.M. Silva, H.P. Rocha, G.M. Almeida, A.P. Lemos, and A.P. Braga. Milkde: A new approach for multiple instance learning based on positive instance selection and kernel density estimation. *Engineering Applications of Artificial Intelligence*, 59:196-204, 2017.

We addressed the issues and shortcomings of existing multi-instance SVM methods. Specifically, we targeted the issues of biased classification caused by the limited available information about instances within positive bags.

- ***Have you published parts of your paper before, for instance in a conference? If so, give details of your previous paper(s) and a precise statement detailing how your paper provides a significant contribution beyond the previous paper(s).***

No, this is a novel contribution that has not been published nor submitted elsewhere in part or its entirety.