Support Vector Machines for Multiple-Instance Learning

Stuart Andrews, Ioannis Tsochantaridis and Thomas Hofmann
Department of Computer Science, Brown University, Providence, RI 02912 {stu,it,th}@cs.brown.edu

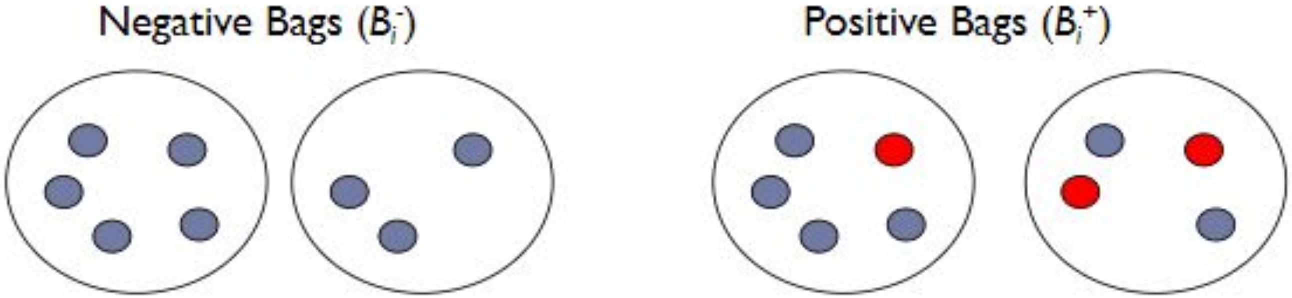
Yawer abbas
B20CI050

What we are doing?

We are using two different model in machine learning and using a heuristic approach to optimized the result.

Multiple Instance Learning (MIL)

- In MIL, instead of giving the learner labels for the individual examples, the trainer only labels collections of examples, which are called bags.
- A bag is labeled positive if there is at least one positive example in it
- It is labeled negative if all the examples in it are negative



Single Instance Learning MIL

- SIL-MIL: Single Instance Learning approach
  - Applies bag's label to all instances in the bag
  - A normal SVM is trained on the resulting dataset

minimize:

$$\mathbf{J}(w, b, \xi) = \frac{1}{2}\|w\|^2 + \frac{C}{L}\sum_{X \in \mathcal{X}}\sum_{x \in X}\xi_x$$

subject to:

$$\begin{aligned}
& w\,\phi(x) + b \leq -1 + \xi_x, \;\; \forall x \in \tilde{\mathcal{X}}_n \\
& w\,\phi(x) + b \geq +1 - \xi_x, \;\; \forall x \in \tilde{\mathcal{X}}_p \quad (*) \\
& \xi_x \geq 0
\end{aligned}$$

Implementation

```
initialize y_i = Y_I for i ∈ I
REPEAT
   compute SVM solution w, b for data set with imputed labels
   compute outputs f_i = ⟨w, x_i⟩ + b for all x_i in positive bags
   set y_i = sgn(f_i) for every i ∈ I, Y_I = 1
   FOR (every positive bag B_I)
      IF (∑_{i∈I} (1 + y_i)/2 == 0)
         compute i* = arg max_{i∈I} f_i
         set y_{i*} = 1
      END
   END
WHILE (imputed labels have changed)
OUTPUT (w, b)
```

```
initialize x_I = ∑_{i∈I} x_i/|I| for every positive bag B_I
REPEAT
   compute QP solution w, b for data set with
      positive examples {x_I : Y_I = 1}
   compute outputs f_i = ⟨w, x_i⟩ + b for all x_i in positive bags
   set x_I = x_{s(I)}, s(I) = arg max_{i∈I} f_i for every I, Y_I = 1
WHILE (selector variables s(I) have changed)
OUTPUT (w, b)
```

Results

For image dataset

Data Set Category	Dims inst/feat	EM-DD	mi-SVM			MI-SVM		
			linear	poly	rbf	linear	poly	rbf
TST1	3224/6668	85.8	93.6	92.5	90.4	**93.9**	93.8	93.7
TST2	3344/6842	84.0	78.2	75.9	74.3	**84.5**	84.4	76.4
TST3	3246/6568	69.0	**87.0**	83.3	69.0	82.2	85.1	77.4
TST4	3391/6626	80.5	82.8	80.0	69.6	82.4	**82.9**	77.3
TST7	3367/7037	75.4	**81.3**	78.7	81.3	78.0	78.7	64.5
TST9	3300/6982	65.5	**67.5**	65.6	55.2	60.2	63.7	57.0
TST10	3453/7073	78.5	79.6	78.3	52.6	79.5	**81.0**	69.1

Classification accuracy of different methods on the TREC9

We have presented a novel approach to multiple-instance learning based on two alternative generalizations of the maximum margin idea used in SVM classification. Although these formulations lead to hard mixed integer problems, even simple local optimization heuristics already yield quite competitive results compared to the baseline approach. We conjecture that better optimization techniques, that can for example avoid unfavorable local minima, may further improve the classification accuracy.

Reference

[1] P. Auer. On learning from multi-instance examples: Empirical evaluation of a theoretical approach. In Proc. 14th International Conf. on Machine Learning, pages 21-29. Morgan Kaufmann, San Francisco, CA, 1997.
[2] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In Proceedings Third International Conference on Visual Information Systems. Springer, 1999.
[3] A. Demirez and K. Bennett. Optimization approaches to semisupervised learning. In M. Ferris, O. Mangasarian, and J. Pang, editors, Applications and Algorithms of Complementarity. Kluwer Academic Publishers, Boston, 2000.
[4] T . G. Dietterich, R. H. Lathrop, and T . Lozano-Perez. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89(1-2):31- 71 , 1997.
[5] T. Gartner, P. A. Flach, A. Kowalczyk, and A. J. Smola. Multi-instance kernels. In Proc. 19th International Conf. on Machine Learning. Morgan Kaufmann, San Francisco, CA, 2002.

[6] T. Joachims. Transductive inference for text classification using support vector machines. In Proceedings 16th International Conference on Machine Learning, pages 200- 209. Morgan Kaufmann, San Francisco, CA, 1999.
[7] P.M. Long and L. Tan. PAC learning axis aligned rectangles with respect to product distributions from multiple-instance examples. In Proc. Compo Learning Theory, 1996.
[8] O. Maron and T. Lozano-Perez. A framework for multiple-instance learning. In Advances in Neural Information Processing Systems, volume 10. MIT Press, 1998.
[9] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In Proc. 15th International Conf. on Machine Learning, pages 341- 349. Morgan Kaufmann, San Francisco, CA, 1998.
[10] J. Ramon and L. De Raedt. Multi instance neural networks. In Proceedings of ICML2000, Workshop on Attribute- Valu e and Relational Learning, 2000.
[11] B. SchOlkopf and A. Smola. Learning with Kernels. Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, 2002.
[12] Qi Zhang and Sally A. Goldman. EM-DD: An improved multiple-instance learning technique. In Advances in Neural Information Processing Systems, volume 14. MIT Press, 2002.