MANUSCRIPT NUMBER: PR-D-17-01043

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

AUTHORS: Gabriella Melki, Alberto Cano, and Sebastián Ventura

Dear Editor and Reviewers,

According to your suggestions, we have prepared a new version of our paper, introducing the changes required to incorporate all the reviewers' comments. The following is our explanation of the changes made, as well as how they address the comments raised by the reviewers. Text in bold corresponds to the original comments. We hope to satisfy the reviewers' expectations. Finally, we are really grateful to the reviewers for their insightful and constructive comments and suggestions in their revisions, which have helped us to improve the quality of the article.

We hope that the reviewers find the proposed thorough revision of our manuscript satisfactory. We look forward to hearing from you soon.

1   Reviewer #1

The authors propose MIRSVM, a support vector machine (SVM) based approach for classification which uses the bag-level information for training. The proposed algorithm is able to identify the bag representatives which are the instances that impact the classification the most. One of the strengths of their method is that it deals with the potential class imbalance by incorporating both positive and negative bags and limiting them to have maximum 1 representative. Unlike other methods the MIRSVM doesn't have any assumption about the data distribution. I support publication of this paper in this journal because:

The topic is very well-suited to the interests of this Journal. The paper is very organized and written smoothly. Enough background information is given about SVM, multi-instance learning and bag level classification. A comprehensive literature review is provided. Each method is shortly described and the pitfalls are discussed which takes the reader to the idea of why an approach like MIRSVM is needed.

We appreciate the reivewer's comments on the background we have given to present our contribution.

The algorithm is clearly described through flowcharts, pseudocodes and examples. All notations are given at the beginning of the paper which helps the reader to follow the equations easier.

We tried our best to provide a comprehensive and detailed explanation of our contribution, and we thank the reviewers for their positive comments on this.

This method is very well evaluated by comparison of the results vs. 11 competing techniques on 15 datasets which provides enough data to the solidity of the evaluations. The authors also have visual comparisons. For example they compare MIRSVM vs MISVM one of the competing techniques which are trained using the same parameters (Fig. 4). The plots clearly indicate the optimal number of support vectors as well as smoother data representation when using MIRSVM. The authors also showed that their method runs faster than all 11 approaches. The statistical analysis and tests are described in detail. By using both precision and recall rates they demonstrate that unlike other methods, MIRSVM doesn't have bias toward positive or negative bags and has balance of precision-recall trade-off. They also perform a statistical analysis using Cohen's Kappa Rate which again proves that MIRSMV outperforms the other methods.

I would strongly recommend this paper to be published in Elsevier, Pattern Recognition Journal.

Again, we would like to express our gratitude on the reviewer's positive comments about our contribution and experimental study. We wanted to be as clear as possible when highlighting the differences between our method and the state-of-the-art methods, and Figure 4 was one of the ways to show how selecting representatives from one class and not the other can prove to be biased. So, we thank the reviewer on aknowledging this.

2   Reviewer #2

The paper presents a new approach for solving multiple-instance classification problem from bag-level perspective. The authors formulated an SVM classifier such that it finds optimal decision hyperplane that separates positive bags from the negative ones. The key idea is to train the SVM using representative instances from each of the bags. The representative instances are those having the largest magnitude of the dot product with the weight vector. In other words, the most positive instances are chosen from positive bags while the least negative instances are chosen from the negative bags. The authors claimed that the approach helps alleviate the data imbalance problem often encountered in multiple-instance learning. The experimental results demonstrated that the proposed method is superior to existing algorithms designed for multiple-instance learning. The proposed method is also faster than many of the existing approaches.

We thank and appreciate the reviewer for the positive and constructive comments given towards improving this paper.

Comment 1: First of all, what I like most about the paper is the way the comparisons are conducted. Not many of the papers that I read/reviewed provided such detailed statistics. Please keep up the good practice.

We tried to be as thorough as possible during our experimental study and we intend on continuing with this method of experimentation. We really appreciate your comment on this.

Comment 2: However, according to me, I think the technical novelty is somewhat marginal. The formulations in Section 3 are all standard SVM, except that it now operates at the bag-level. The only (technical) contribution I see is the new measure used to select bag representatives. And this contribution should be highlighted.

We took your advice on this and removed some of the standard SVM formulations in Section 3 (specifically the Langrangian and derivation of the dual), highlighted the bag-representative selector, as well as emphasized the major benefits using it in conjunction with the bag level formulation of the SVM. Therefore, this way we enhance and clarify the benefits of the contribution, without having redundant information within the section.

Comment 3: I also feel that Section 1 and 2 are too lengthy. It feels like reading a book chapter on multiple-instance learning. These two parts combined are almost half of the manuscript. Also many of the details in Section 3 could be left out as they are not much different from the standard SVM. That is to say, I would like the authors to consider making those parts more concise.

We rewrote and reorganized Sections 1, 2, and 3 to condense the information given. We shortened and rewrote some of the descriptions on the background of multiple-instance learning, as well as removed some of the basic SVM formulations from the background to avoid repetition in Section 3. Your comment on the length of those sections makes sense, thank you.

Comment 4: Lastly, I think the standard MI assumption only makes sense in some problem domains. (Dietterich's drug-activity prediction seems to be the case). So, I believe it would make the paper stronger if the authors also present some interesting real-world usage of the proposed algorithm.

Thank you for noting this. As you have suggested, we added a couple of additional applications of the standard MI assumption to the paper. Specifically, the following real-world scenarios provide a more global picture as to why MI algorithms are useful: bioinformatics (protein identification), image classification & retrieval, web mining & text classification, and medical diagnosis & imaging.

Speicific Comments:

1. *I think your ranges for C parameter and kernel width are too narrow, especially on the lower end. Is there any specific reason for the choice apart from to speed up the cross validation process? Besides, I usually see people using the range 2^i where $i = -10 : 10$. So, how about considering some smaller values as well, say* 0.00001 *and/or* 0.001?

   According to your suggestion, we tested the performance of our method using a wider range of parameters (especially with lower C values). Results were exactly the same as originally reported, since the best

parameters were still selected in the domain we originally defined. The only difference of evaluating a larger number of parameters would be a longer runtime, but we observed no performance difference in the results for any of the datasets evaluated. Thank you.

2. **Page 5, Line 8: "they assume the distribution the instances in positive bags is positive", I don't understand this sentence, please clarify. What does 'distribution is positive' mean in this context?**

   Thank you for pointing out that statement. What we meant was that they assume that all the instances within positive bags are assumed to be positive, or to follow the distribution of positively labelled data. This statement was clarified in the text.

3. **Page 12, first sentence: "This formulation...", what formulation? I suppose some equations are missing?**

   We apologize for the ambiguity. What we were specifying was the MIRSVM bag-representative selector and formulation, referenced by Equations 12 and 13. We have made the appropriate change in text to clarify the meaning.

4. **Page 12, Line 5: "... where each bag is allowed one." $\rightarrow$ I don't think this is grammatical.**

   We've updated the text to the correct grammar. Thank you.

5. **Algorithm 1 Line 8: Graham matrix $\rightarrow$ Gram matrix?**

   Thank you for making note of that. We have updated the text to be "Gram matrix".

Thank you for your comprehensive notes and constructive comments on our paper.