OXFORD

Data and text mining

RMTL: an R library for multi-task learning

Han Cao[1], Jiayu Zhou[2] and Emanuel Schwarz[1,*]

[1]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, 68159 Mannheim, Germany and [2]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Multi-task learning (MTL) is a machine learning technique for simultaneous learning of multiple related classification or regression tasks. Despite its increasing popularity, MTL algorithms are currently not available in the widely used software environment R, creating a bottleneck for their application in biomedical research.

Results: We developed an efficient, easy-to-use R library for MTL (www.r-project.org) comprising 10 algorithms applicable for regression, classification, joint predictor selection, task clustering, low-rank learning and incorporation of biological networks. We demonstrate the utility of the algorithms using simulated data.

Availability and implementation: The RMTL package is an open source R package and is freely available at https://github.com/transbioZl/RMTL. RMTL will also be available on cran.r-project.org.

Contact: emanuel.schwarz@zi-mannheim.de

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Multi-task learning (MTL) is a machine learning technique that explores and exploits the relatedness across a set of different learning tasks. Since its inception (Caruana, 1998), MTL has been used in numerous data-intensive research areas, including biomedical informatics (Feriante, 2015; Li *et al.*, 2016; Widmer and Ratsch, 2012; Xu *et al.*, 2011; Yuan *et al.*, 2016; Zhou *et al.*, 2013), speech and natural language processing [i.e. (Wu *et al.*, 2015)], image processing and computer vision [i.e. (Wang *et al.*, 2009)], as well as web based applications [i.e. (Chapelle *et al.*, 2010)].

A strong motivation to develop biomedical MTL applications stems from the necessity to integrate diverse data sources to explore the biological underpinning of complex illnesses, such as schizophrenia. Previous research has already shown that for such illnesses, integrative multi-omics open a new avenue for identification of etiological mechanisms, for example by taking into account genetic, expression and methylation data simultaneously [i.e. (Lin *et al.*, 2014)]. For such applications, multi-task learning offers the possibility to directly explore illness-related biological profiles that are linked across data modalities and therefore a new route toward the identification of biomarker signatures.

Previous implementations of MTL have focused on knowledge transfer via regularization (Zhou *et al.*, 2011), Bayesian methods (Greenlaw *et al.*, 2017) or deep architectures (Yang and Hospedales, 2016). Here, we developed the first R library for MTL, offering a comprehensive machine learning pipeline that covers several types of MLT algorithms and can be easily applied to high-dimensional data.

In the following section, we briefly describe the RMTL package, including the implemented MTL methods (for detailed information see Supplementary Methods). The results section describes the application of the algorithms on a simulation study, to demonstrate the performance and interpretability of the respective models.

2 Materials and methods

This package provides an automated, simple-to-use implementation of MTL, comprising five classification and five regression algorithms, which share knowledge across tasks according to different priors via regularization. All algorithms aim to minimize the same objective:

$$\min_{W} \sum_{i}^{t} \frac{1}{n_i} \mathcal{L}(W_i | X_i, \ Y_i) + \Omega(W)$$

where $L(\circ)$ is the loss function (logistic loss for classification or least

square loss for regression). $X = \{X_i\}$, $Y = \{Y_i\}$ are sets of predictor matrices and the corresponding responses for all t tasks where $X_i \in R^{n_i \times p}$ and $Y_i \in R^{n_i \times 1}$ is the predictor matrix and the response vector of task $i \in \{1, 2, \ldots, t\}$. Accordingly, n_i and p refer to the number of subjects and predictors (all tasks share the same predictor space) of task i, respectively. Moreover, $W = R^{p \times t}$ is the coefficient matrix for all tasks, where W_i, the ith column of W, is the coefficient vector for task i.

Knowledge transfer among tasks is achieved via a convex term $\Omega(W)$ that jointly modulates models according to specific functionalities. In this package, five common regularization techniques are implemented to suit different applications, i.e. sparse structure, joint predictor selection, low-rank structure, network constraint for task relatedness and task clustering. Here, we refer to the above regularization strategies as MTL_Lasso, MTL_L21, MTL_Trace, MTL_Graph and MTL_CMTL, in the same sequence. These strategies can be broadly categorized into two classes: strategies for predictor selection (MTL_Lasso and MTL_L21) and strategies for task relatedness exploration (MTL_Graph, MTL_Trace and MTL_CMTL). While the former class explores sparse patterns are explored over the predictor space, the latter class exploits task relatedness based on additional assumptions. For all algorithms, we implemented a solver based on the accelerated gradient descent method (Nesterov, 2013). To solve the non-smooth and convex regularization, the proximal operator (Parikh and Boyd, 2014) was applied. Overall, the solver achieves a complexity of $O(1/k^2)$, which is optimal among first-order gradient methods. Further methodological details are shown in the Supplementary Methods.

3 Results

Predictive performance and model interpretability of the implemented algorithms were explored using simulated data. The simulated datasets were constructed by the ground truth model W, which is specified for a given prior (Supplementary Fig. S1). We compared the ground truth and the learnt model as an indicator of model interpretability. For predictive comparison, the primary baseline method was the conventional lasso, which reflects single task learning performance. We further applied MTL with lasso (MTL_Lasso), to explore the effect of inappropriate prior choice as a second baseline method.

3.1 Model interpretability

Supplementary Figure S1a shows the coefficient matrix of MTL_Lasso and MTL_L21 and demonstrates that the number of predictors identified by MTL_Lasso was approximately half the number of ground truth predictors. This may be due to the fact that highly correlated predictors exist in the high-dimensional space (Zou and Hastie, 2005). As a consequence and similar to conventional Lasso, MTL_Lasso tended to select one among several correlated predictors. Despite this, 75% (precision) of selected predictors were ground truth predictors. For MTC_L21, the ground truth was highly sparse: only 40 out of 400 predictors were active predictors for all tasks. The simulation demonstrates that 39 of the predictors were successfully identified (sensitivity: 97.5%), with a precision of 72%. These results indicate that MTL algorithms could successfully identify ground truth predictors.

The relatedness of tasks was represented by pairwise correlation between models. Supplementary Figure S1b shows that all methods were able to capture correctly the pairwise relatedness compared to the ground truths. Particularly, MTL_Graph incorporated a strong network prior such that the "in-group" differences became zero. This may be because the network prior provided the most complete information about task relatedness among all priors.

3.2 Predictive performance

Supplementary Figure S2 indicates that conventional Lasso failed to yield accurate predictions on all simulated datasets except when using the l_{21} prior. Compared to this baseline, the MTL models improved the accuracy by 18.7% on average. The MTL_Lasso incorporating an inappropriate prior achieved an average accuracy of 67% and was substantially inferior to MTL models with appropriate priors (average accuracy: 79.2%).

4 Conclusion

In this study, we developed an R library for multi-task learning comprising 10 algorithms incorporating five different priors. MTL models outperformed two baseline methods when applied on simulated data. High model-interpretability was observed in terms of predictor selection and task-relatedness compared to the respective ground truths.

Funding

Conflict of Interest: none declared.

References

Caruana,R. (1998) *Multitask Learning*. Springer, USA.

Chapelle,O. *et al*. (2010) Multi-task learning for boosting with application to web search ranking. In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '10*.

Feriante,J. (2015). *Massively Multitask Deep Learning for Drug Discovery*. University of Wisconsin-Madison.

Greenlaw,K. *et al*. (2017) A Bayesian group sparse multi-task regression model for imaging genetics. *Bioinformatics*, **33**, 2513–2522.

Li,Y. *et al*. (2016) A Multi-Task Learning Formulation for Survival Analysis. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowedge Discovery and Data Mining (KDD'16)*.

Lin,D. *et al*. (2014) Integrative analysis of multiple diverse omics datasets by sparse group multitask regression. *Front. Cell Dev. Biol.*, **2**, 62.

Nesterov,Y. (2013) Gradient methods for minimizing composite functions. *Math. Program.*, **140**, 125–161.

Parikh,N., and Boyd,S. (2014) Proximal algorithms. *Found. Trends Optim.*, **1**, 127–239.

Wang,X. *et al*. (2009) Boosted multi-task learning for face verification with applications to web image and video search. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Patter Recognition*.

Widmer,C. and Ratsch,G. (2012) Multitask learning in computational biology. *JMLR*, **27**, 207–216.

Wu,Z. *et al*. (2015) Deep neural networks employing multi-task learning and stacked bottleneck features for speech synthesis. In: *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*.

Xu,Q. *et al*. (2011) Multitask learning for protein subcellular location prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **8**, 748–759.

Yang,Y. and Hospedales,T. (2016) Deep multi-task representation learning: a tensor factorisation approach. *arXiv Preprint arXiv*, **1605**, 06391.

Yuan,H. *et al*. (2016) Multitask learning improves prediction of cancer drug sensitivity. *Sci. Rep.*, **6**, 31619.

Zhou,J. *et al*. (2011) *Malsar: Multi-task Learning via Structural Regularization*. Vol. **21**. Arizona State University.

Zhou,J. *et al*. (2013) Modeling disease progression via multi-task learning. *Neuroimage*, **78**, 233–248.

Zou,H. and Hastie,T. (2005) Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B*, **67**, 301–320.