# Active semi-supervised learning for multi-target regression

Maira Farias Andrade Lira[1], Luisa Cavalcante[1], Celine Vens[2,3], Ricardo Prudencio *[2,3], and Felipe Kenji Nakano[2,3]

[1] Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil
[2] KU Leuven, Campus KULAK, Department of Public Health and Primary Care
[3] Itec, imec research group at KU Leuven
mfal@cin.ufpe.br, lncc@cin.ufpe.br, celine.vens@kuleuven.be, rbcp@cin.ufpe.br, felipekenji.nakano@kuleuven.be

**Abstract.** Recent works have proposed the combination of active and semi-supervised learning techniques to efficiently incorporate unlabeled data. The so-called active semi-supervised learning (ASSL) investigates methods to efficiently construct predictive models by employing either a domain expert or model to (pseudo)-label data. To the best of our knowledge, ASSL has not been applied in the context of multi-target regression, a predictive task where multiple continuous targets must be predicted. In this work, we propose MASSTER, Multi-target Active Semi-Supervised Training for Regression, a novel ensemble method that identifies the most relevant instance-target pairs based on the variance in their predictions. Further, as its semi-supervised component, our method incorporates a variation of both self-learning (MASSTER-SL) and co-training (MASSTER-CT). Experiments using 8 benchmark datasets reveal that our method provides superior results in most of the cases when compared to the current state-of-the-art.

**Keywords:** Active learning · Semi-supervised learning · Multi-target regression

## 1 Introduction and Method

Supervised machine learning algorithms often require substantial amounts of labeled data to perform well, which is not always available. In many cases, however, unlabeled data is easily obtainable. In order to reduce labeling costs, studies have proposed active semi-supervised learning (ASSL), the combination of active learning (AL) with semi-supervised learning (SSL), which incorporates data labeled by an oracle and data is pseudo-labelled using a model.

To the best of our knowledge, ASSL has never been investigated in multi-target regression (MTR), a predictive task where multiple continuous outputs must be predicted. We propose MASSTER, Multi-target Active Semi-Supervised

* These authors have equally contributed to this manuscript.

Training for Regression, a method that uses the variance of predictions to identify the most relevant instance-target pairs. Further, it is combined with both self-learning [6] (MASSTER-SL) and co-training [3] (MASSTER-CT), well-established SSL methods.

By using 8 benchmark datasets, we have compared the proposed method with both (1) the only AL method for MTR [1] [2] and (2) adaptations of AL methods originally proposed for single-target regression [7] [5], including the popular method greedy sampling [7] and RT-AL, the current state-of-the-art [5].

## 2 Experiments

For a complete analysis of the results, please refer to our paper [4]. In Figure 1, we present our results. As can be seen, the variants of our proposed method, MASSTER-CT and MASSTER-SL, provided superior in most of the cases, where their curves rise faster than the AL counterpart (MASSTER-AL) and the SSL counterparts (SSL-SL and SSL-CT). Thus, our methods are capable of identifying the most informative data points while correctly pseudo-labeling part of the unlabeled pool.

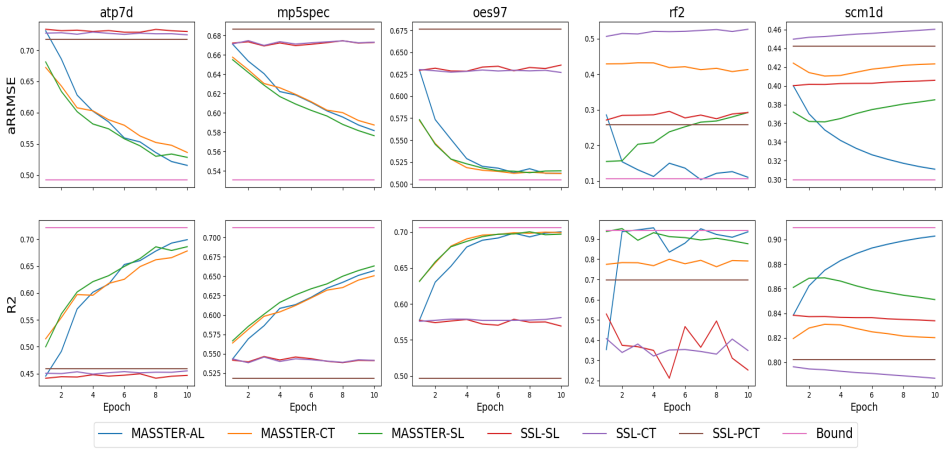


Fig. 1: The x-axis contains the epochs, whereas the y-axis shows the performance.

## 3 Conclusion

To the best of our knowledge, we have presented the first work on semi-supervised active learning for multi-target regression. More specifically, we proposed a method that employs the variance of predictions, along with a semi-supervised component. Future work should include novel active learning algorithms for multi-target regression.

## References


1. Quinten Danneels, Felipe Kenji Nakano, and Celine Vens. Active learning for multi-target regression. Master's thesis, KU Leuven, 2022.
2. Thomas Dirkse, A. J Feelders, M. M. Dastani, and J. M. van den Akker. Active learning for multi-target regression problems, with an application to meta-modeling of transportation simulations. Master's thesis, Utrecht University, 2019.
3. Jun Du, Charles X Ling, and Zhi-Hua Zhou. When does cotraining work in real data? *IEEE Transactions on Knowledge and Data Engineering*, 23(5):788–799, 2010.
4. Maira Farias Andrade Lira, Luisa Cavalcante, Celine Vens, Ricardo Prudencio, and Felipe Kenji Nakano. Active semi-supervised learning for multi-target regression. In *2025 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2025.
5. Ashna Jose, João Paulo Almeida de Mendonça, Emilie Devijver, Noël Jakse, Valérie Monbet, and Roberta Poloni. Regression tree-based active learning. *Data Mining and Knowledge Discovery*, 38(2):420–460, 2024.
6. Isaac Triguero, Salvador García, and Francisco Herrera. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. *Knowledge and Information systems*, 42:245–284, 2015.
7. Dongrui Wu, Chin-Teng Lin, and Jian Huang. Active learning for regression using greedy sampling. *Information Sciences*, 474:90–105, 2019.