SINGLE-CELL ADAPTIVE RESAMPLING (SC-AR)

OVERVIEW

sc-AR is a repository of sample code for an adaptive resampling approach to enhance representation learning of single-cell RNA-sequencing data.

WHAT CAN SC-AR DO

This repository was developed to implement an adaptive resampling (AR) approach to resampling single-cell data based on its learned latent structure in an online, adaptive manner. The repository provides sample code to implement and evaluate this approach on single-cell RNA-sequencing datasets.

INTENDED USES

The sc-AR repository is suited for computational biologists and bioinformaticians interested in single-cell transcriptomic datasets. It is being shared with the research community to facilitate reproduction of our results and foster further research in this area. This repository intended to be used by domain experts who are independently capable of evaluating the quality of outputs before acting on them.

OUT-OF-SCOPE USES

sc-AR is not well suited for tasks outside of single-cell biology. It is an algorithm to improve machine learning in single-cell transcriptomic and genomic datasets.

sc-AR does not involve any LLM or generative AI models, and does not possess any related abilities. sc-AR is not intended for use in the context of high-risk decision making or in sensitive domains.

HOW TO GET STARTED

To begin using sc-AR, please follow the instructions in our repository README: https://github.com/microsoft/sc-AR.

EVALUATION

The adaptive resampling algorithm was evaluated on its ability to improve gene expression reconstruction, cell type classification, and perturbation response in single-cell RNA-seq.

A detailed discussion of our evaluation methods and results can be found in our paper, linked in the sc-AR repository.

EVALUATION METHODS

We used a variety of metrics and downstream tasks to measure the performance of AR in machine learning of scRNA-seq.

We compared the performance of AR against standard training algorithms on the gene expression reconstruction, cell type classification, and perturbation response scRNA-seq tasks.

The models used for evaluation were autoencoder based. Additional details can be found in our preprint manuscript. Future work on additional model architectures will help test AR further; results may vary if sc-AR is used with different models.

MICROSOFT RESPONSIBLE AI TRANSPARENCY DOCUMENTATION FOR RESEARCH

EVALUATION RESULTS

We found that AR leads to significantly improved downstream performance across datasets and metrics. Additionally, it enhances the quality of learned cellular embeddings compared to standard training methods. Our results suggest that AR may serve as a valuable technique for improving representation learning and predictive performance in single-cell transcriptomic models.

LIMITATIONS

sc-AR was developed for research and experimental purposes, and is intended as a research tool.

BEST PRACTICES

Details are provided in the repository README.

LICENSE

MIT License

TRADEMARKS

This project may contain trademarks or logos for projects, products, or services. Authorized use of Microsoft trademarks or logos is subject to and must follow Microsoft's Trademark & Brand Guidelines. Use of Microsoft trademarks or logos in modified versions of this project must not cause confusion or imply Microsoft sponsorship. Any use of third-party trademarks or logos are subject to those third-party's policies.

CONTACT

This research was conducted by members of Microsoft Research, former Microsoft Research interns, and of the Broad Institute of MIT & Harvard. We welcome feedback and collaboration from our audience. If you have suggestions, questions, or observe unexpected/offensive behavior in our technology, please contact Ava Amini at ava.amini@microsoft.com.