OXFOR

Doc-Start

Subject Section

koios: machine learning for big biological data

Simon Dirmeier [1] and Niko Beerenwinkel [1],*

[1]Department, Institution, City, Post Code, Switzerland

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Abstract

Motivation: Analysing biological data is becoming increasingly difficult due to its ever increasing volumnes and dimensionality. Since desktop solutions are not sufficient anymore to analyse big biological data sets comprising terabytes of data, such as features extracted from images or gene expression of millions of cells, we propose a novel machine learning command line tool called *koios* for a distributed highly-parallel analysis of biological data.

Availability and implementation: *koios* is available from Github (https://github.com/cbg-ethz/koios) and soon to be released on Bioconda.

Contact: niko.beerenwinkel@bsse.ethz.ch

1 Introduction

The not-so-recent advent of high-dimensional data is still posing not only statistical, but primarily methodological problems for researchers across fields such as computatopnal biology, astronomy or the social sciences. (1) typisches gelaber: recent advent of high blablabla

2 Methods

dimensionality reduction example Factor analysis clustering regression/prediction/random forests

example regression

3 Conclusion

Acknowledgements

Text Text Text Text Text Text Text Text. **?** might want to know about text text text text

References

[1] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.