Cornell University Library

We gratefully acknowledge support from
the Simons Foundation
and member institutions

Ensembling Classification Models Based on Phalanxes of Variables with Applications in Drug Discovery

Jabed H Tomal, William J Welch, Ruben H Zamar

(Submitted on 20 Mar 2013 (v1), last revised 27 Apr 2013 (this version, v2))

We have proposed an ensemble method which aggregates over clusters of predictor variables. We form the clusters (we call phalanxes) by joining variables together. The variables in a phalanx are good to put together, and the variables in different phalanxes are good to ensemble. We then build our ensemble of phalanxes (EPX) by growing a random forest (RF) in each phalanx and aggregating them over the phalanxes. We have applied our ensemble EPX to rank rare active compounds ahead of the majority inactive compounds in four drug discovery assay datasets, and compared its performances with random forest and regularized random forest (RRF). Five descriptor sets are tried for each of the four assays. Our ensemble EPX was found superior to RF and RRF in most of the times. The improvement of our ensemble over RF and RRF is impressive when datasets contain a small proportion of active compounds and/or many predictors.

Comments:	Withdrawn without any reason
Subjects:	**Machine Learning (stat.ML)**; Computation (stat.CO)
Cite as:	**arXiv:1303.4805 [stat.ML]**
	(or **arXiv:1303.4805v2 [stat.ML]** for this version)

Submission history

From: Jabed Hossain Tomal [view email]
[v1] Wed, 20 Mar 2013 01:23:50 GMT (67kb)
[v2] Sat, 27 Apr 2013 18:13:27 GMT (0kb,I)

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.

Download:

- Source

Current browse context:

stat.ML

< prev | next >

new | recent | 1303

Change to browse by:

stat

stat.CO

References & Citations

- NASA ADS

Bookmark(what is this?)

Science WISE