Data Mining Lab: Dynamic Weighted Majority for Incremental Learning

Miena Basta Badres, David Jozefiak

Introduction
Overview

- Results of the paper vs. reproduced results
- Performance of DWMIL with F1-score
- Performance of DWMIL on a new data set
- DWMIL vs. Learn^{++}.NIE

Results of the paper vs. reproduced results
Geometric Mean

data set	paper	reproduced results
Moving Gaussian	0.7565	0.9115
SEA	0.9256	0.7705
Hyper Plane	0.5889	0.6460
Checkerboard	0.8123	0.6508
Electricity	0.7062	0.7550
Weather	0.6641	0.7405

Results of the paper vs. reproduced results
AUC

data set	paper	reproduced results
Moving Gaussian	0.8517	0.9115
SEA	0.9776	0.7705
Hyper Plane	0.7007	0.6466
Checkerboard	0.8876	0.6510
Electricity	0.8271	0.7557
Weather	0.7725	0.7408

Performance of DWMIL with F1-score

data set	f1-score
Moving Gaussian	0.9116
SEA	0.8155
Hyper Plane	0.6612
Checkerboard	0.6571
Electricity	0.7197
Weather	0.6417

Performance of DWMIL on new data sets
Forest Covertype

- DWMIL performs very good on this real-world data set

metric	value
gm	0.9222
f1	0.8040
auc	0.9211
rec	0.9129

Learn++.NIE
Competitor

- Learn++ for Non-stationary and Imbalanced Environments
- modified algorithm of Learn++.CDS
    - employs a different penalty constraint that forces the algorithm to balance predictive accuracy on all classes
    - uses a bagging based sub-ensemble for the minority class oversampling

For more details

Ditzler, Gregory & Polikar, Robi. (2013). Incremental Learning of Concept Drift from Streaming Imbalanced Data. Knowledge and Data Engineering, IEEE Transactions on. 25. 2283-2301. 10.1109/TKDE.2012.136.

DWMIL vs. Learn⁺⁺.NIE
Forest Covertype

- DWMIL performs better at this real-world data set

metric	DWMIL	Learn⁺⁺.NIE
gm	0.9222	0.8984
f1	0.8040	0.7838
auc	0.9211	0.8928
rec	0.9129	0.8616

DWMIL vs. Learn++.NIE
Moving Gaussian

- both methods perform equally good

metric	DWMIL	Learn++.NIE
gm	0.9115	0.9282
f1	0.9116	0.9291
auc	0.9115	0.9281
rec	0.9114	0.9252

DWMIL vs. Learn++.NIE
SEA

- Learn++.NIE performs better at this data set

metric	DWMIL	Learn++.NIE
gm	0.7705	0.9231
f1	0.8155	0.9350
auc	0.7705	0.9217
rec	0.7848	0.9503

DWMIL vs. Learn⁺⁺.NIE
Hyper Plane

- Learn⁺⁺.NIE performs better at this data set

metric	DWMIL	Learn⁺⁺.NIE
gm	0.6460	0.8464
f1	0.6612	0.8558
auc	0.6466	0.8474
rec	0.6224	0.8263

DWMIL vs. Learn⁺⁺.NIE
Checkerboard

- Learn⁺⁺.NIE performs better at this data set

metric	DWMIL	Learn⁺⁺.NIE
gm	0.6508	0.8631
f1	0.6571	0.8633
auc	0.6510	0.8626
rec	0.6374	0.8793

DWMIL vs. Learn⁺⁺.NIE
Electricity

- both methods perform equally good

metric	DWMIL	Learn⁺⁺.NIE
gm	0.7550	0.7415
f1	0.7197	0.6988
auc	0.7557	0.7489
rec	0.7241	0.6474

DWMIL vs. Learn++.NIE
Weather

- both methods perform equally good

metric	DWMIL	Learn++.NIE
gm	0.7405	0.7247
f1	0.6417	0.6307
auc	0.7408	0.7312
rec	0.7617	0.8375