DriftSurf: Stable-State / Reactive-State Learning under Concept Drift

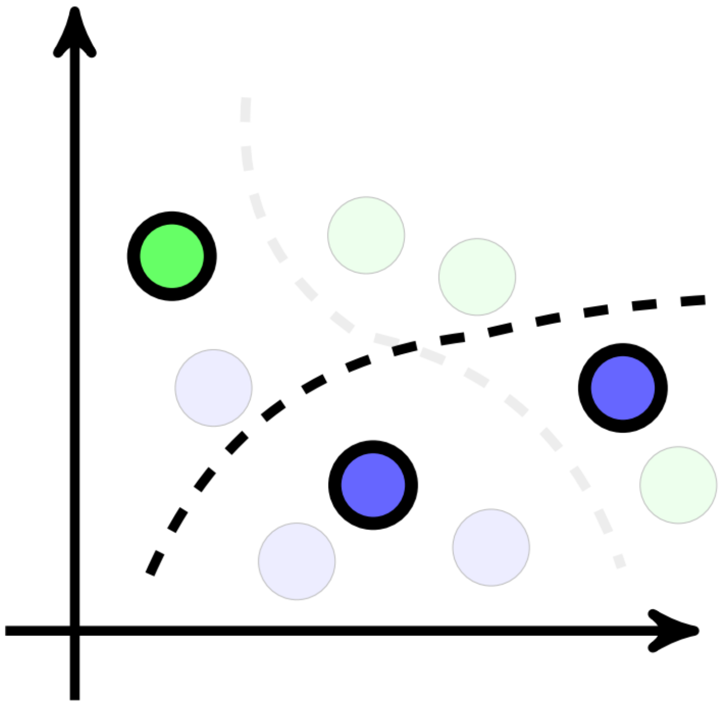
Ashraf Tahmasbi
Srikanta Tirthapura
(Iowa State)

Ellango Jothimurugesan
Phillip B. Gibbons
(CMU)

Problem Definition

- At each time step $t = 1, 2, \ldots,$ a batch of data points arrive
- Data at time t are sampled from a distribution $P_t(X, y)$
- Concept drift occurs at time t if $P_t(X, y) \neq P_{t-1}(X, y)$
- Objective at each time t is to minimize the expected risk:

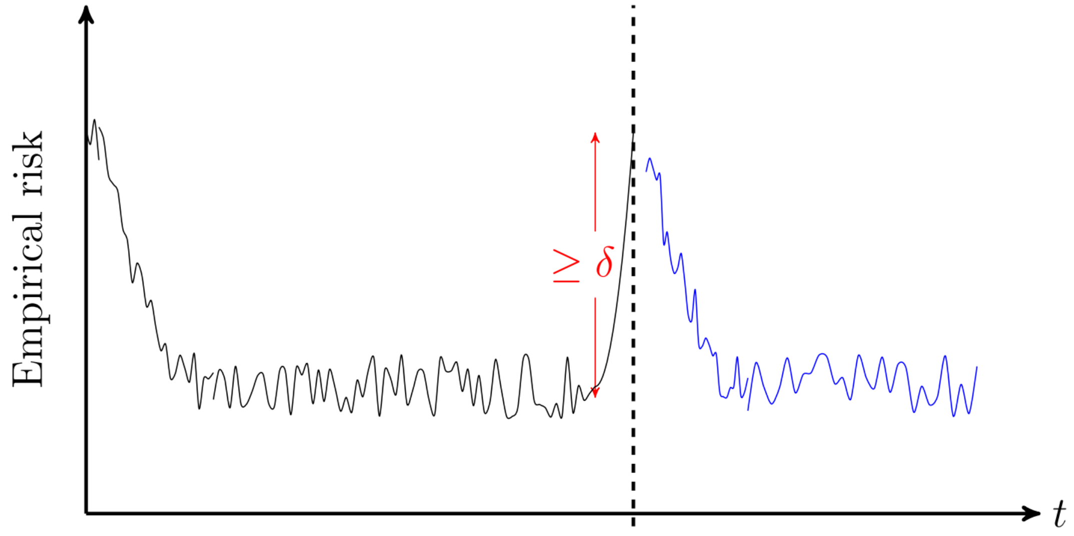
$$\min_{w_t \in \mathcal{F}} \mathbb{E}_{(X,y)\sim P_t} \left[\ell(w_t(X), y)\right]$$



Shortcomings of Prior Drift Detection

Test: do the model's empirical risks degrade over time by a set threshold δ?

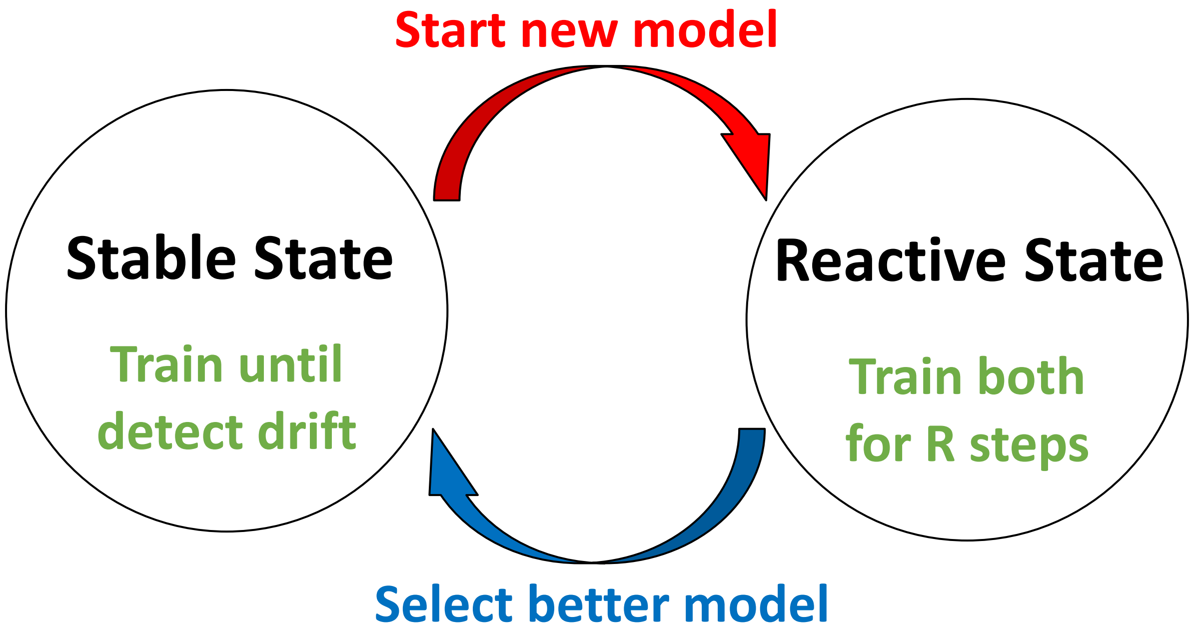
E.g., PERM [Harel et al. ‘14], HDDM [Frías-Blanco et al. ‘15], MDDM [Pesaranghader et al. ‘18]



Successful detection followed by model replacement

- x Difficult to tune threshold, trading false negatives and false positives
- x False negatives: miss subtle drifts
- x False positives: discard a long-trained model

DriftSurf Improves on Prior Drift Detection



Further details in paper:

- Detection test uses frozen models for slow drifts
- Case of drift during reactive state

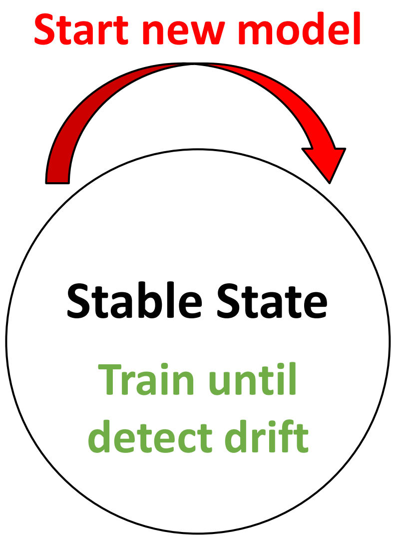
✓Apply aggressive detection (small threshold) to catch even subtle drifts

✓Mitigate false positives and keep the long-trained model

During reactive state, use best predictor from previous step

✓Reacts quickly to true drifts and accounts for model warm-up

Summary of Theoretical Results (I)



- **StandardDD** is solely the drift detection test used in DriftSurf

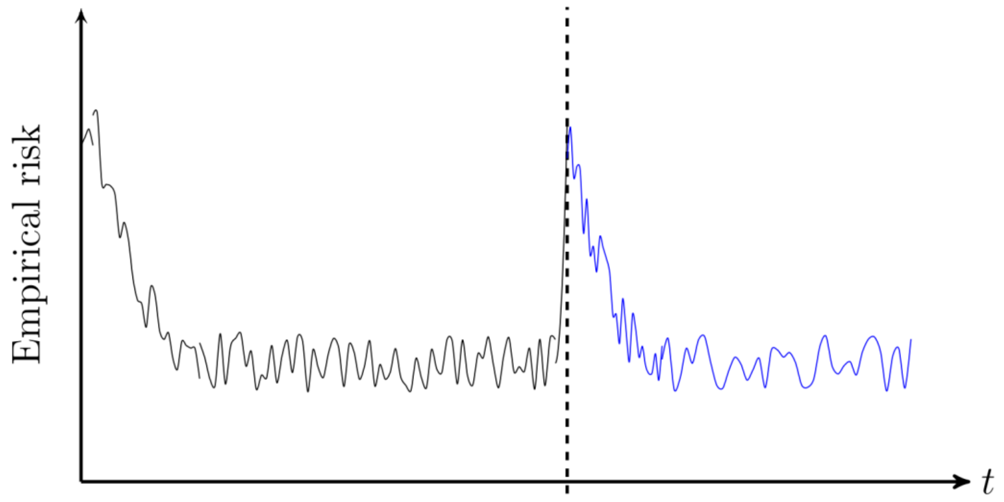
- *DriftSurf has higher statistical accuracy than StandardDD*

  In the absence of drifts, and as $t \to \infty$:
  - Theorem: Expected age of DriftSurf's model is $\geq t/4$
  - Best known bound: Expected age of StandardDD's model is bounded by a constant

DriftSurf models are long-trained

Summary of Theoretical Results (II)

- We show DriftSurf's time to recover from a drift is small
- **Aware** is an idealized algorithm with oracle access to when drifts occur



Aware resets the model at the time of drift

- *DriftSurf is risk-competitive to Aware*
  - Theorem: W.h.p., at each time step after DriftSurf recovers, the expected risk bound of DriftSurf is a constant factor of Aware's

Experimental Results

Drift detection methods:

- MDDM [Pesaranghader et al. '18]
- StandardDD

Ensemble method:

- AUE [Brzezinski & Stefanowski '13]

Misclassification rates averaged over time
(base learner: logistic regression)

ALGORITHM / DATASET	AUE	MDDM	STANDARDDD	DriftSurf	Aware
SEA0	0.093	**0.086**	0.097	**0.086**	0.137
SEA20	0.245	0.289	0.249	**0.243**	0.264
SEA-GRADUAL	0.162	0.165	0.160	**0.159**	0.177
HYPER-SLOW	**0.112**	0.116	0.116	0.118	0.110
HYPER-FAST	0.179	**0.163**	0.168	0.173	0.191
SINE1	0.212	**0.176**	0.184	0.187	0.171
MIXED	0.209	**0.204**	**0.204**	**0.204**	0.192
CIRCLES	0.379	0.372	0.377	**0.371**	0.368
RCV	0.167	**0.125**	0.126	**0.125**	0.121
COVERTYPE	0.279	0.311	**0.267**	0.268	0.267
AIRLINE	**0.333**	0.345	0.338	0.334	0.338
ELECTRICITY	0.296	0.344	0.320	**0.290**	0.315
POWERSUPPLY	0.301	0.322	0.308	**0.292**	0.309

Accuracy Comparison

- High accuracy for both gradual & abrupt drifts
- Outperform AUE and StandardDD
- Generally outperform MDDM, especially on real datasets

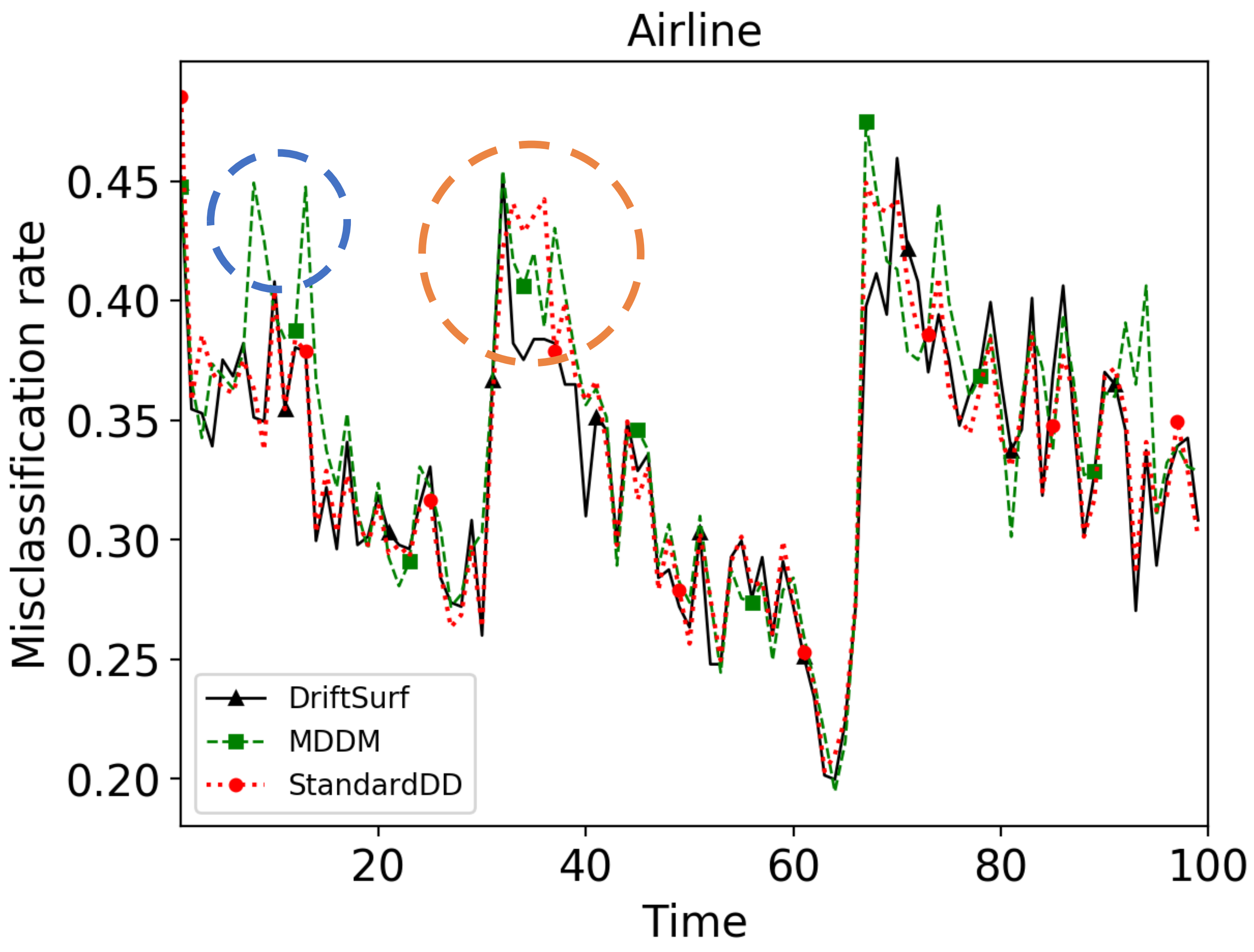
Misclassification rates averaged over time
(base learner: logistic regression)

Algorithm / **Dataset**	AUE	MDDM	Stand-ardDD	DriftSurf	Aware
SEA0	0.093	**0.086**	0.097	**0.086**	0.137
SEA20	0.245	0.289	0.249	**0.243**	0.264
SEA-Gradual	0.162	0.165	0.160	**0.159**	0.177
Hyper-slow	**0.112**	0.116	0.116	0.118	0.110
Hyper-fast	0.179	**0.163**	0.168	0.173	0.191
SINE1	0.212	**0.176**	0.184	0.187	0.171
Mixed	0.209	**0.204**	**0.204**	**0.204**	0.192
Circles	0.379	0.372	0.377	**0.371**	0.368
RCV	0.167	**0.125**	0.126	**0.125**	0.121
CoverType	0.279	0.311	**0.267**	0.268	0.267
Airline	**0.333**	0.345	0.338	0.334	0.338
Electricity	0.296	0.344	0.320	**0.290**	0.315
PowerSupply	0.301	0.322	0.308	**0.292**	0.309

Accuracy on Airline Dataset over Time

- MDDM suffers from false positives

- StandardDD and MDDM perform worse after drift



Robustness to Drift Detection Threshold

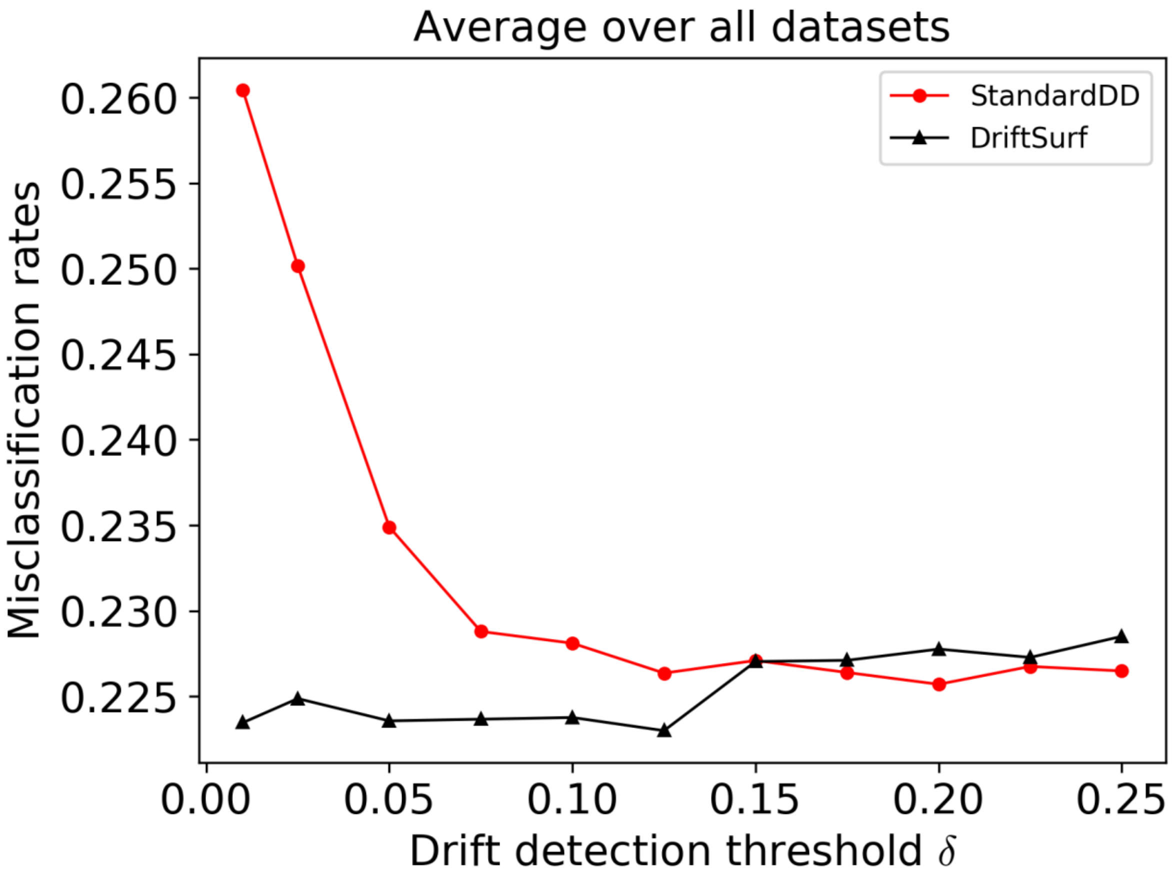
StandardDD:

x Difficult to tune threshold

DriftSurf:

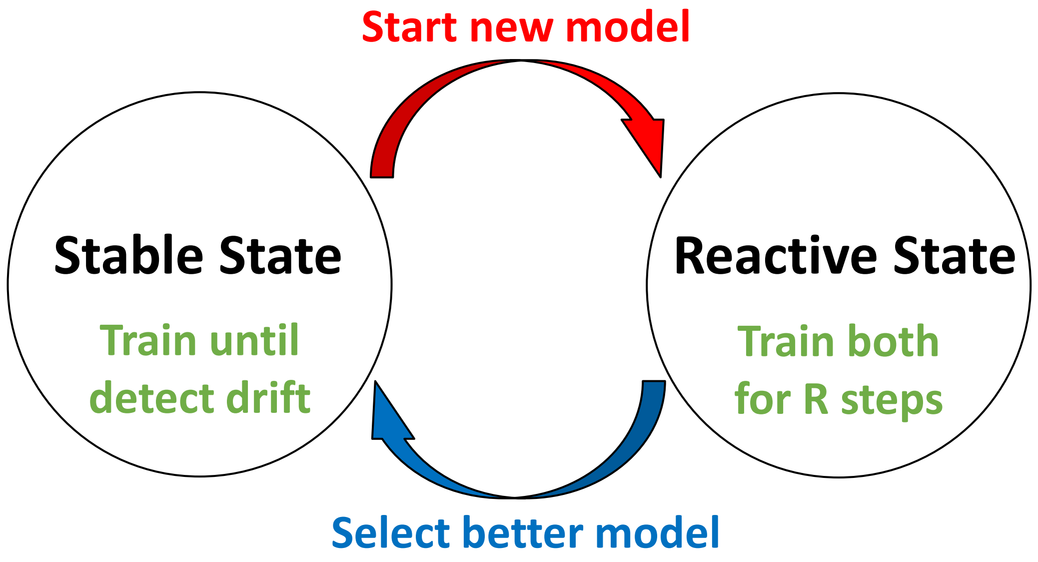
✓ Apply aggressive detection (small threshold) to catch even subtle drifts

✓ Mitigate false positives and keep the long-trained model



aggressive conservative

Conclusion



- DriftSurf is risk-competitive to Aware (oracle knowledge of drifts)
- DriftSurf's reactive-state approach provides statistically better learning than standalone drift detection
- Our experimental evaluation shows high accuracy in the presence of abrupt and gradual drifts