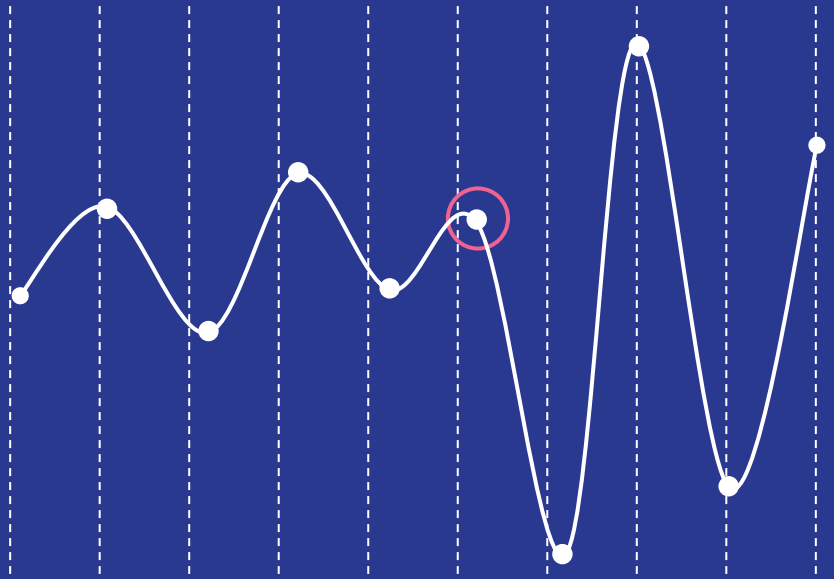
Algorithm Recommendation in Data Streams

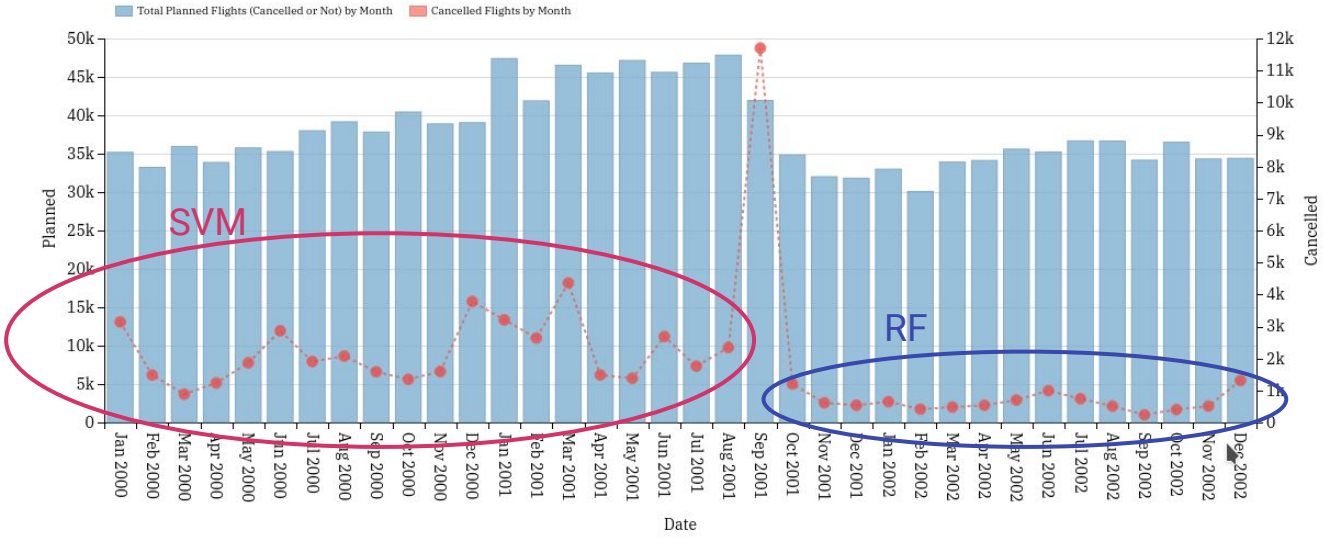
Meta-Learning and Concept Drift



Motivation

Cancelled Flights from/to NY in 2000 - 2002

The impact of terrorist attacks (11-september) on US flights.



Data Streams

AWS Spot Price

```
2017-05-08 21:46:36+00:00 c3.8xlarge Windows ap-northeast-1c 1.7461
2017-05-08 21:46:34+00:00 i3.large SUSE Linux ap-northeast-1c 0.1223
2017-05-08 21:46:34+00:00 i3.large Linux/UNIX ap-northeast-1c 0.0223
2017-05-08 21:46:17+00:00 c4.8xlarge SUSE Linux ap-northeast-1a 0.789
2017-05-08 21:46:17+00:00 c4.8xlarge Linux/UNIX ap-northeast-1a 0.689
2017-05-08 21:46:17+00:00 m2.4xlarge SUSE Linux ap-northeast-1c 0.2782
2017-05-08 21:46:17+00:00 m2.4xlarge Linux/UNIX ap-northeast-1c 0.1782
2017-05-08 21:46:10+00:00 r3.2xlarge SUSE Linux ap-northeast-1c 0.2282
2017-05-08 21:46:10+00:00 r3.2xlarge Linux/UNIX ap-northeast-1c 0.1282
2017-05-08 21:46:09+00:00 r3.xlarge SUSE Linux ap-northeast-1c 0.1536
2017-05-08 21:46:09+00:00 r3.xlarge Linux/UNIX ap-northeast-1c 0.0536
2017-05-08 21:46:05+00:00 i2.xlarge SUSE Linux ap-northeast-1a 0.2134
2017-05-08 21:46:05+00:00 i2.xlarge Linux/UNIX ap-northeast-1a 0.1134
2017-05-08 21:46:04+00:00 c4.large SUSE Linux ap-northeast-1c 0.13
2017-05-08 21:46:04+00:00 c4.large Linux/UNIX ap-northeast-1c 0.03
2017-05-08 21:46:03+00:00 c4.8xlarge Windows ap-northeast-1a 1.6561
2017-05-08 21:46:03+00:00 c4.xlarge SUSE Linux ap-northeast-1c 0.1749
2017-05-08 21:46:03+00:00 c4.xlarge Linux/UNIX ap-northeast-1c 0.0749
2017-05-08 21:46:01+00:00 m1.xlarge SUSE Linux ap-northeast-1c 0.141
2017-05-08 21:46:01+00:00 m1.xlarge Linux/UNIX ap-northeast-1c 0.041
2017-05-08 21:46:00+00:00 m3.xlarge SUSE Linux ap-northeast-1c 0.156
.
.
.
```

S is a Data Stream

S = {a1, a2, a3, …}

ax, instances of this stream

a1 = (s1, Δ1)

sx, index

Δx, describers the instance

Issues

- $|\mathbf{S}| = \infty$;
- **S** may present concept drift;

Method

ELSEVIER

Neurocomputing
Volume 127, 15 March 2014, Pages 52-64

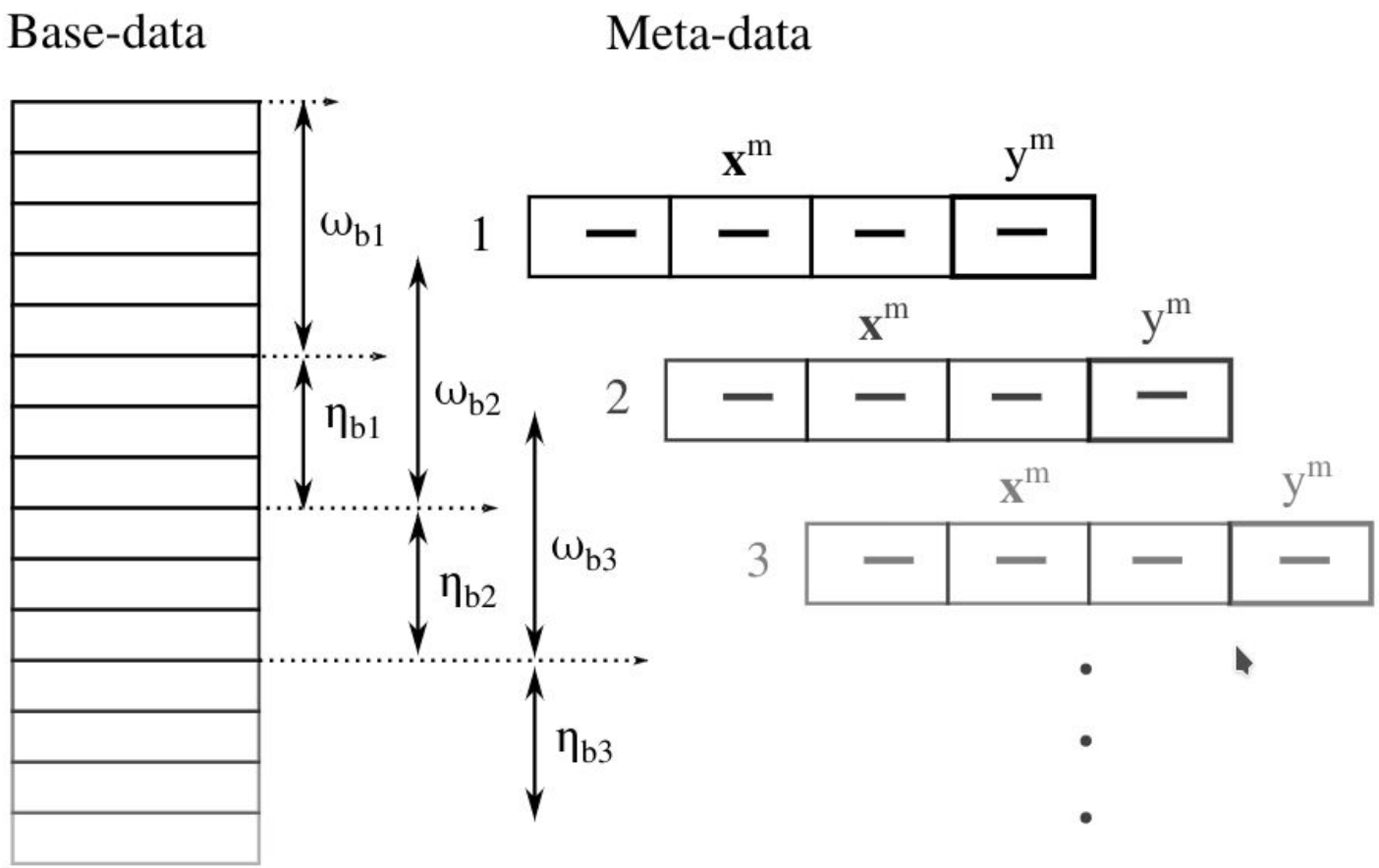
MetaStream: A meta-learning based method for periodic algorithm selection in time-changing data

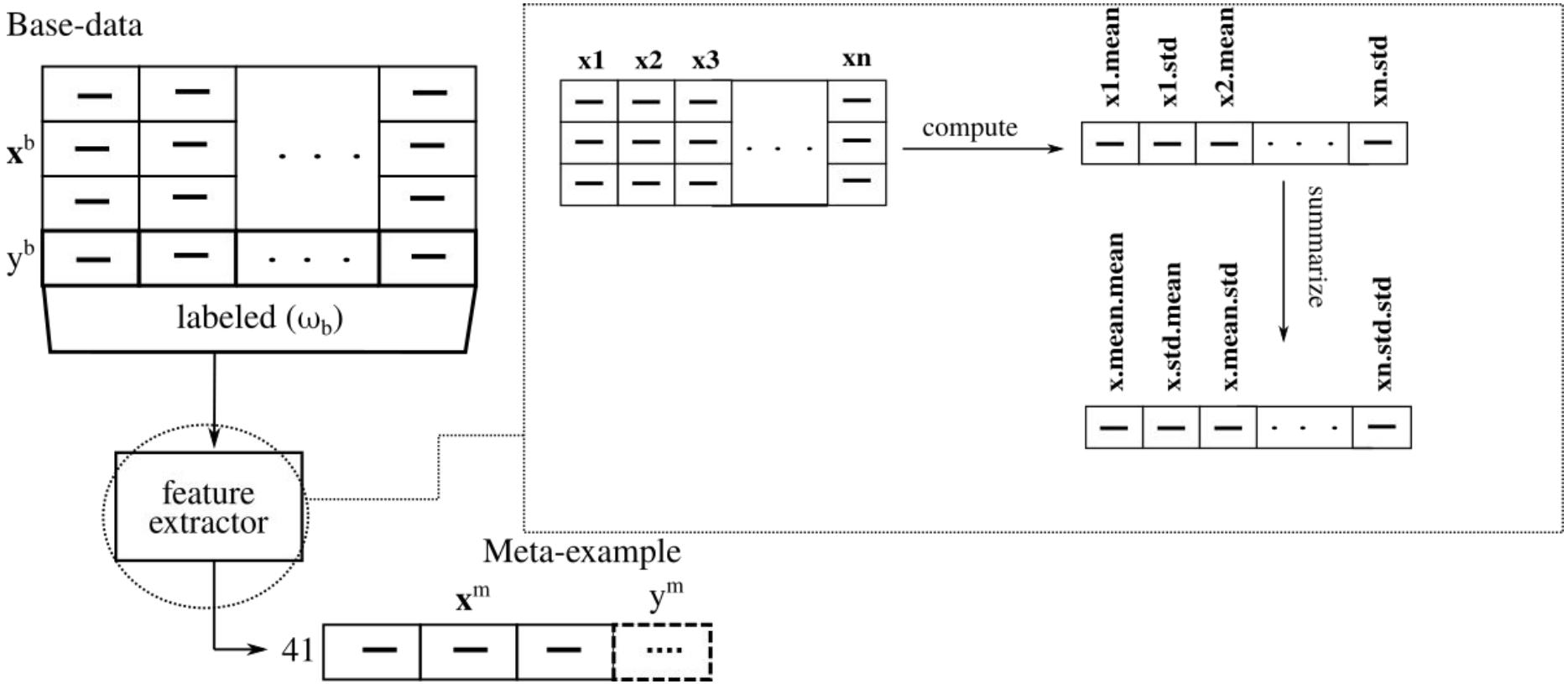
André Luis Debiaso Rossi [a], André Carlos Ponce de Leon Ferreira de Carvalho [a], Carlos Soares [b], Bruno Feres de Souza [a]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.neucom.2013.05.048


Base-data
Meta-data
ω_{b1}
η_{b1}
ω_{b2}
η_{b2}
ω_{b3}
η_{b3}
1
2
3
$\mathbf{x}^m$
y^m


Base-data
$\mathbf{x}^b$
y^b
labeled (ω_b)
feature extractor
x1
x2
x3
xn
compute
x1.mean
x1.std
x2.mean
xn.std
summarize
x.mean.mean
x.std.mean
x.mean.std
xn.std.std
Meta-example
$\mathbf{x}^m$
y^m
41

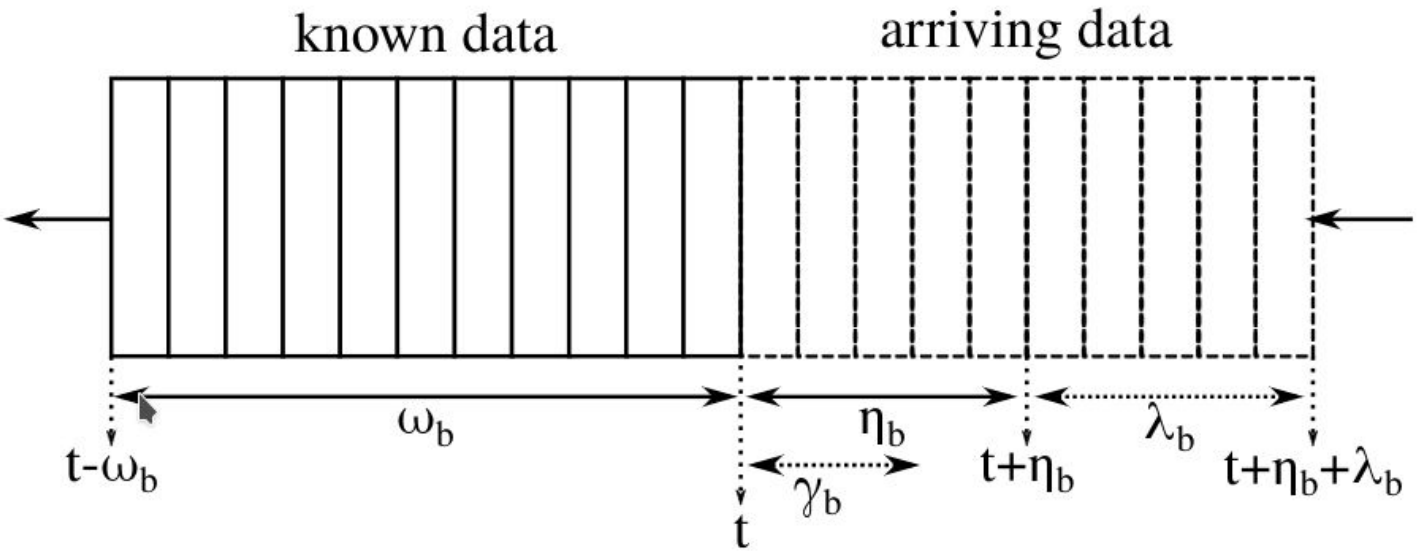
Original vs Improvement

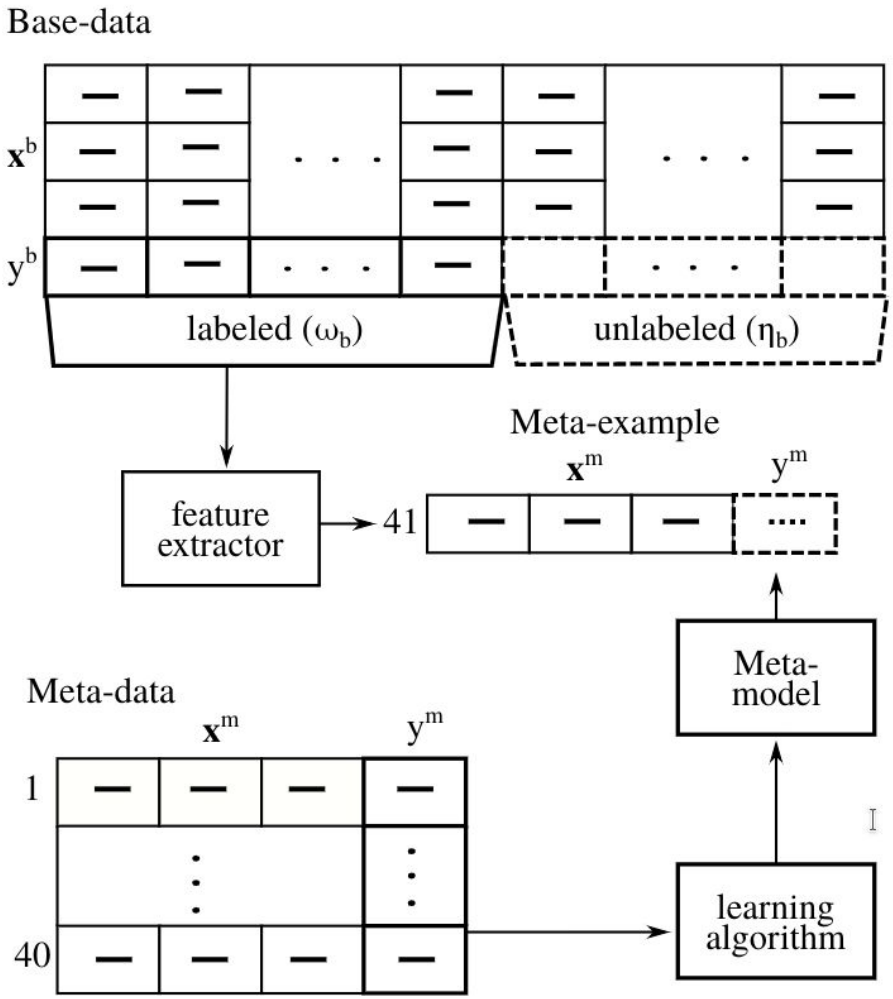
- Statistical
  - Average
  - Variance
  - Minimum
  - Maximum
  - Median
  - Correlation
  - Correlation of numeric attributes to the target
  - Possibility of existence of outliers in numeric attributes
  - Possibility of existence of outliers in the target
  - Dispersion gain
  - Skewness
  - Kurtosis

- Statistical
  - can_cor, cov, g_mean, gravity, h_mean, iq_range, lh_trace, mad, sparsity, t_mean, w_lambda, nr_cor_attr, nr_disc, nr_normnr_outliers, p_trace, range, roy_root
- Landmarking
  - best_node, elite_nn, linear_discr, naive_bayes, one_nn, random_node, worst_node
- Model-based
  - leaves, leaves_branch, leaves_corrob, leaves_homo, leaves_per_class, nodes, nodes_per_attr, nodes_per_inst, nodes_per_level, nodes_repeated, tree_depth, tree_imbalance, tree_shape, var_importance

LightGBM

- Automatically select best features
- Low memory usage
- Good empirical results
- Incremental learning (lower memory and processing time)
- Automatically deals with missing features (present in meta-data)


known data
arriving data
ω_b
η_b
λ_b
$t-\omega_b$
γ_b
$t+\eta_b$
$t+\eta_b+\lambda_b$
t


Base-data
$\mathbf{x}^b$
y^b
labeled (ω_b)
unlabeled (η_b)
feature extractor
41
Meta-example
$\mathbf{x}^m$
y^m
Meta-model
Meta-data
$\mathbf{x}^m$
y^m
1
40
learning algorithm

Settings

Classifiers: [SVM, Random Forest]

Windows: Train 300, Test 30, Step 30 (base and meta levels)

Incremental and Non-Incremental

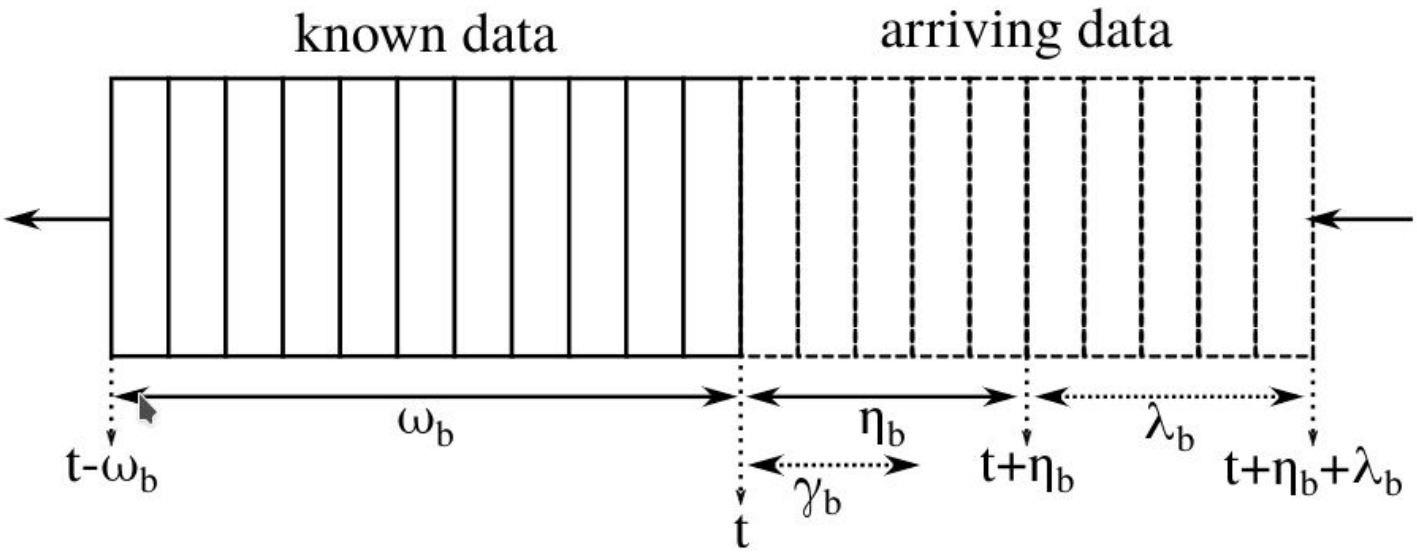

known data
arriving data
ω_b
η_b
λ_b
$t-\omega_b$
γ_b
$t+\eta_b$
$t+\eta_b+\lambda_b$
t

Table 1: Distribution between algorithms for datasets.

Conjunto de Dados	SVM	RF
Electricity	0.752	0.248
CoverType	0.718	0.282
PowerSupply	0.535	0.465
HyperPlane	0.792	0.207
Agrawal	0.785	0.215
RandomRBF	0.535	0.465

Results

Offline

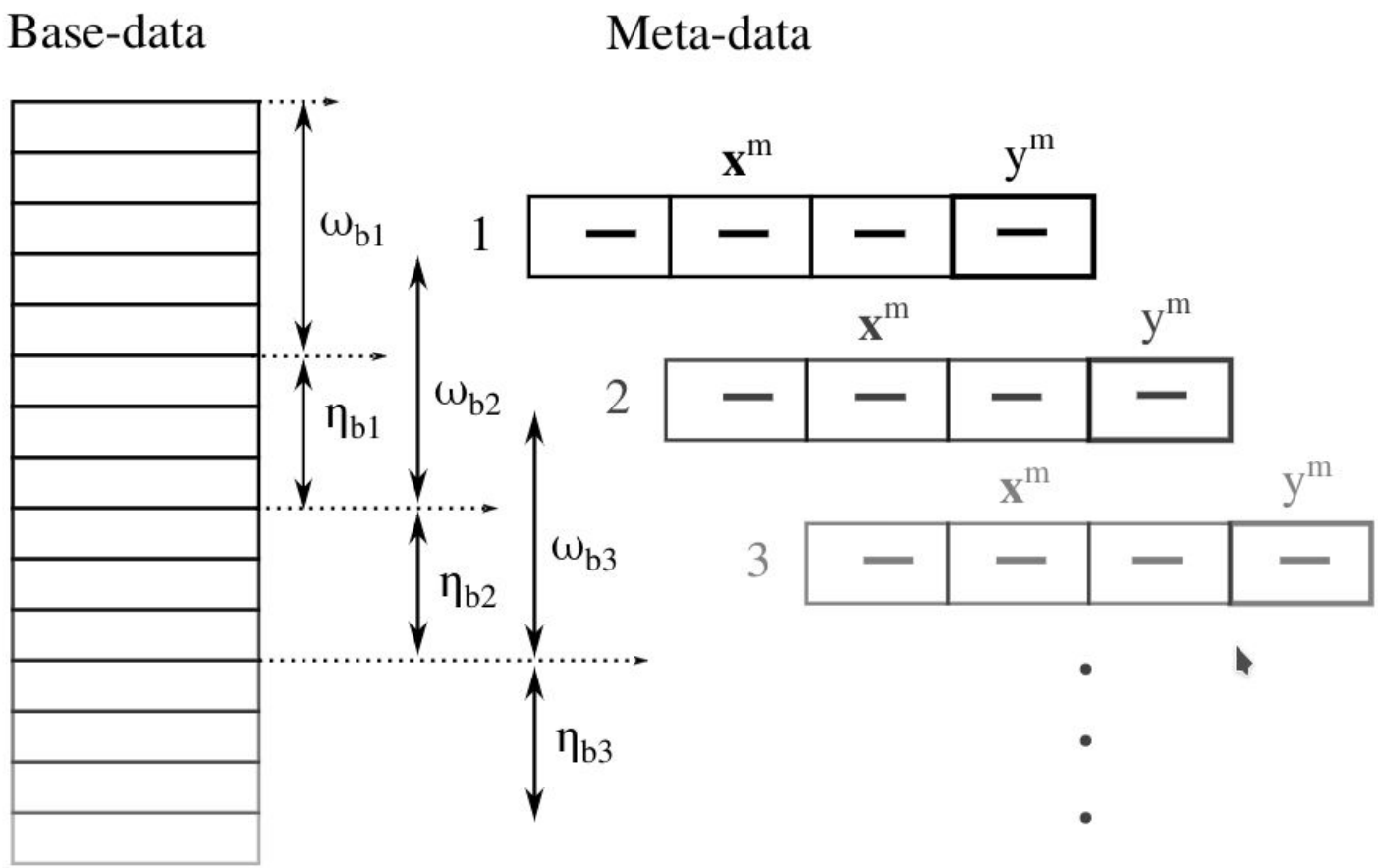
Table 2: Meta-classifier evaluation in offline experiment.

Fluxo	Kappa	M. Geométrica	Acurácia
Electricity	0.095±0.176	0.366±0.207	0.610±0.152
CoverType	0.018±0.215	0.329±0.232	0.517±0.147
PowerSupply	0.251±0.282	0.520±0.252	0.611±0.188
HyperPlane	0.118±0.326	0.120±0.325	0.841±0.080
Agrawal	0.007±0.138	0.046±0.164	0.780±0.078
RandomRBF	0.120±0.239	0.416±0.262	0.571±0.115

Online

Table 3: Meta-classifier evaluation in online experiment.

Dataset	Estratégia	Kappa	M. Geométrica	Acurácia
Electricity	Non-Incremental	0.027	0.462	0.558
	Incremental	0.039	0.422	0.457
CoverType	Non-Incremental	0.114	0.461	0.694
	Incremental	-0.009	0.405	0.631
PowerSupply	Non-Incremental	0.092	0.502	0.607
	Incremental	0.074	0.541	0.539
HyperPlane	Non-Incremental	0.014	0.265	0.747
	Incremental	0.020	0.283	0.745
Agrawal	Non-Incremental	0.010	0.198	0.794
	Incremental	-0.041	0I327	0.700
RandomRBF	Non-Incremental	-0.031	0.484	0.484
	Incremental	0.004	0.486	0.501


Base-data
Meta-data
ω_{b1}
η_{b1}
ω_{b2}
η_{b2}
ω_{b3}
η_{b3}
1
2
3
$\mathbf{x}^m$
y^m

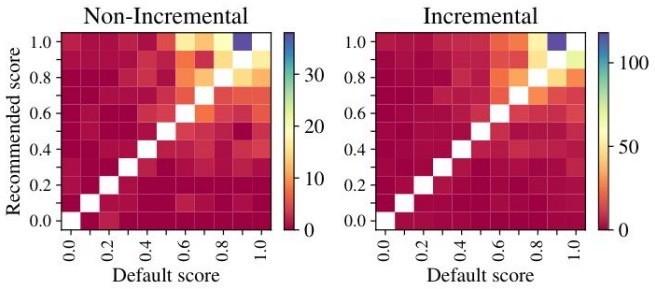


Figura 4.8: Comparison between recommended and Default for Electricity.

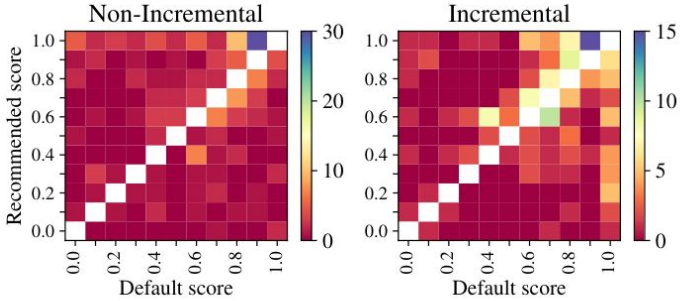


Figura 4.9: Comparison between recommended and Default for CoverType.

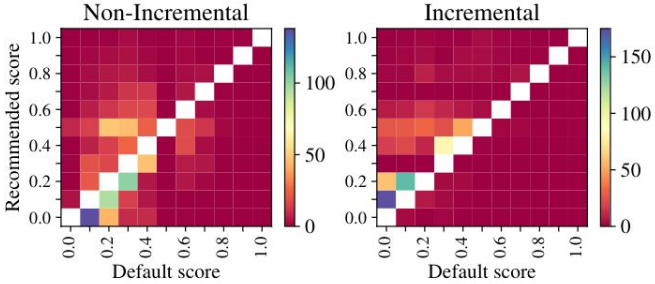


Figura 4.10: Comparison between recommended and Default for PowerSupply.

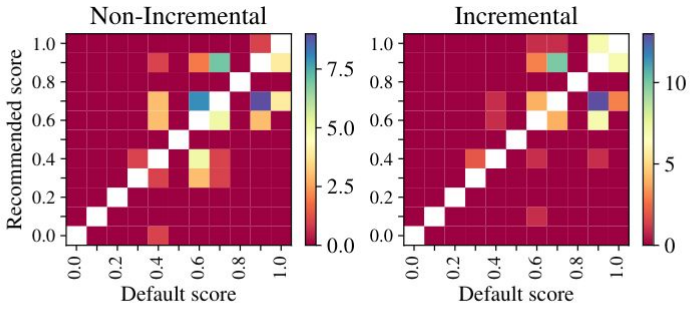


Figura 4.11: Comparison between recommended and Default for HyperPlane.

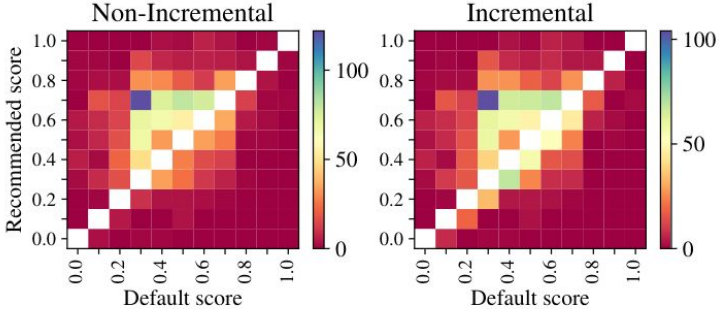


Figura 4.12: Comparison between recommended and Default for Agrawal.

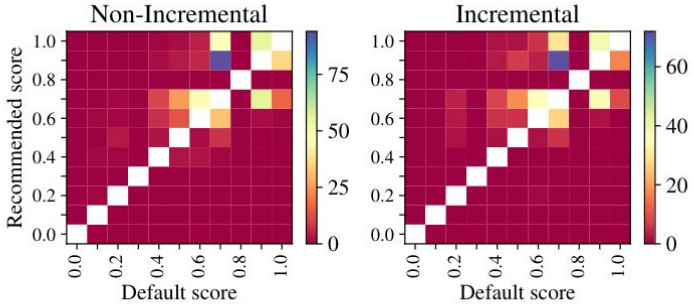


Figura 4.13: Comparison between recommended and Default for RandomRBF.

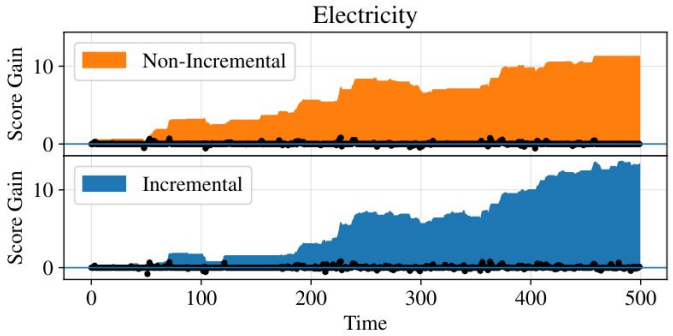


Figura 4.2: Ganho cumulativo de escore ao longo do tempo para o conjunto Electricity.

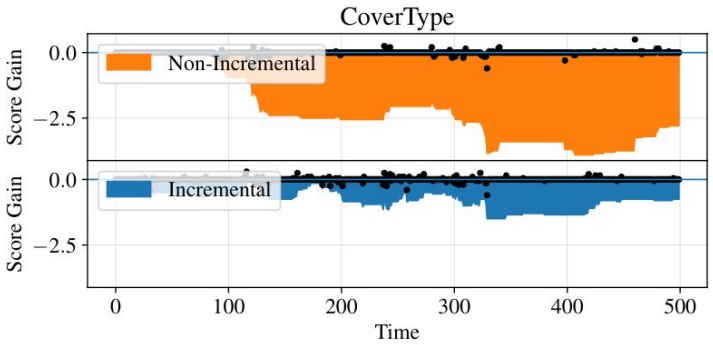


Figura 4.3: Ganho cumulativo de escore ao longo do tempo para o conjunto CoverType.

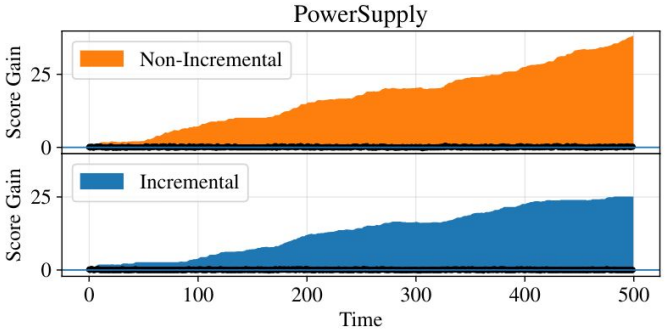


Figura 4.4: Ganho cumulativo de escore ao longo do tempo para o conjunto PowerSupply.

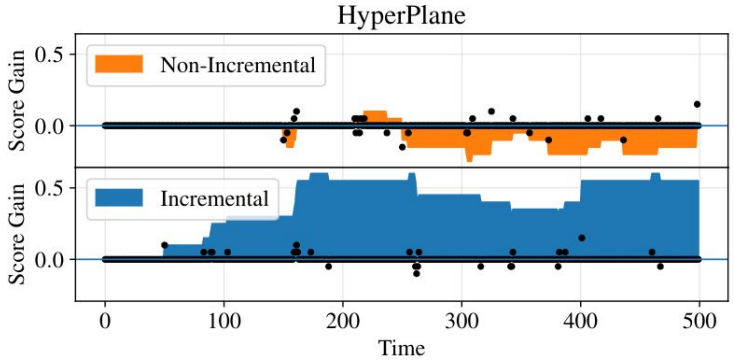


Figura 4.5: Ganho cumulativo de escore ao longo do tempo para o conjunto HyperPlane.

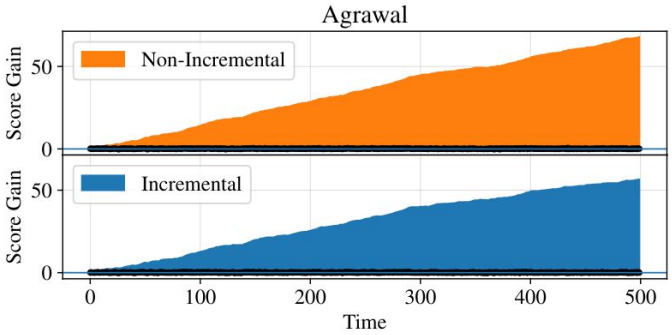


Figura 4.6: Ganho cumulativo de escore ao longo do tempo para o conjunto Agrawal.

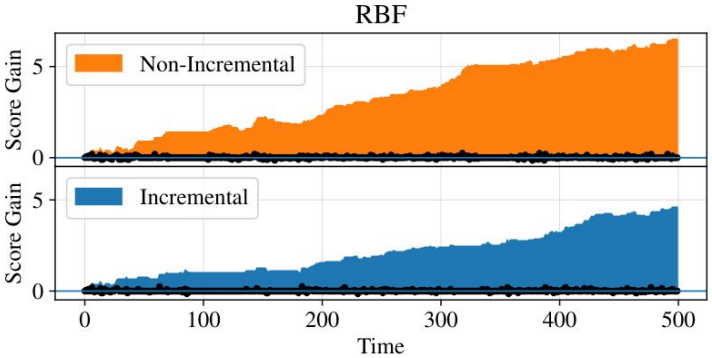


Figura 4.7: Ganho cumulativo de escore ao longo do tempo para o conjunto RandomRBF.

Conclusion

The end