Meta-learning for Periodic Algorithm Selection in Time-changing Data

André L.D. Rossi
André C.P.L.F. de Carvalho
Carlos Soares

ICMC USP São Carlos
Universidade do Porto FEUP Faculdade de Engenharia

Masaryk U. – Advanced Machine Learning
based on a presentation given at the 2012 Brazilian Symposium on Neural Networks

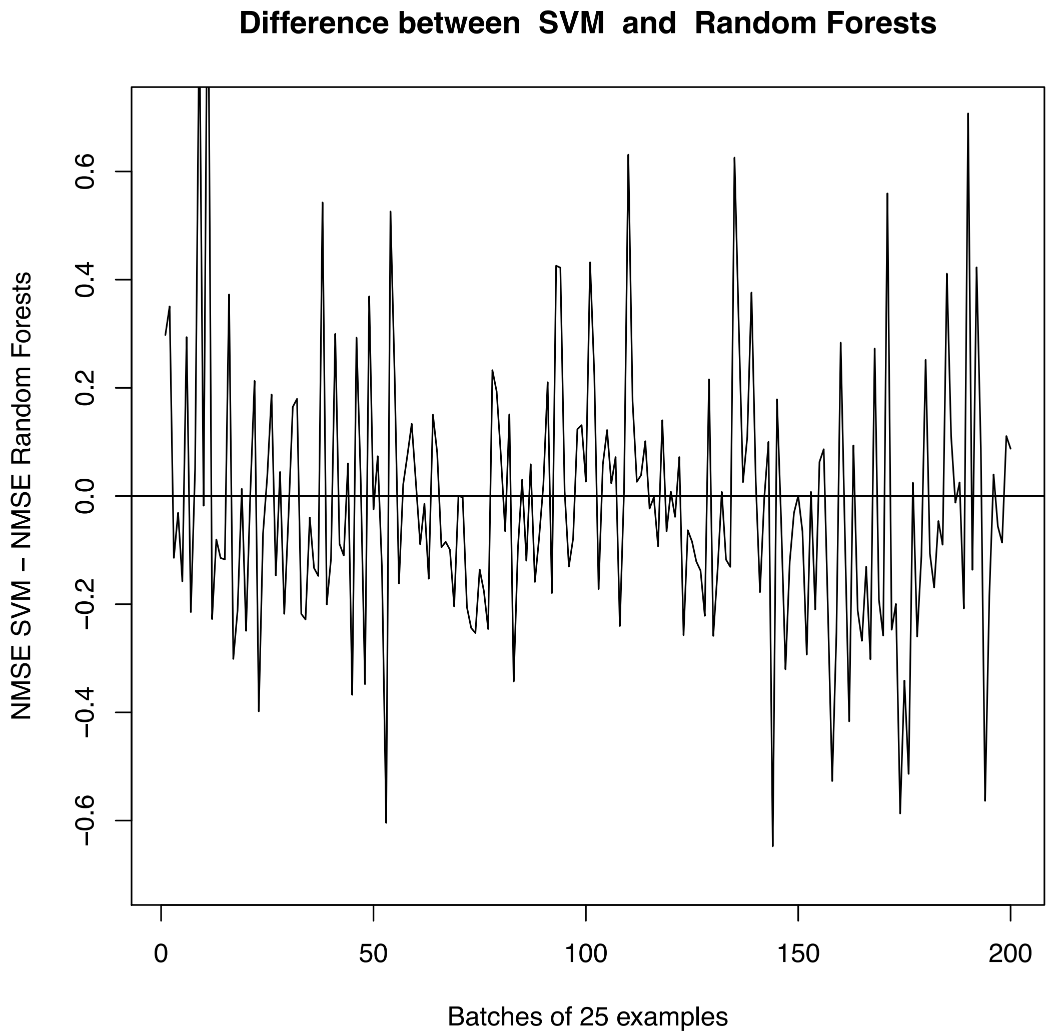
Outline

- Introduction
- MetaStream
- Experiments
- Results
- Meta-data issues
- Conclusions and Future Work

Context

- Dynamic real-world systems generate data continuously
  - Underlying distribution naturally changes over time
    - Concept drift: gradual or abrupt
- Different performance of learning algorithms for different instants of time
- No-free Lunch Theorem (Wolpert, 1996)
  - Any learning algorithm cannot always be the best for all possible learning tasks

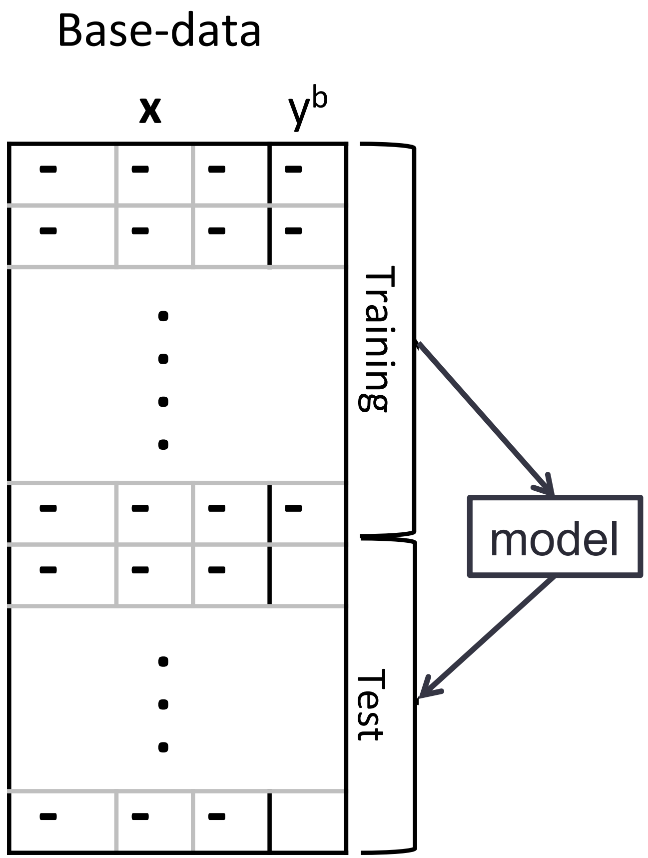
Motivation


Difference between SVM and Random Forests
NMSE SVM – NMSE Random Forests
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0
50
100
150
200
Batches of 25 examples

Problem & Approach

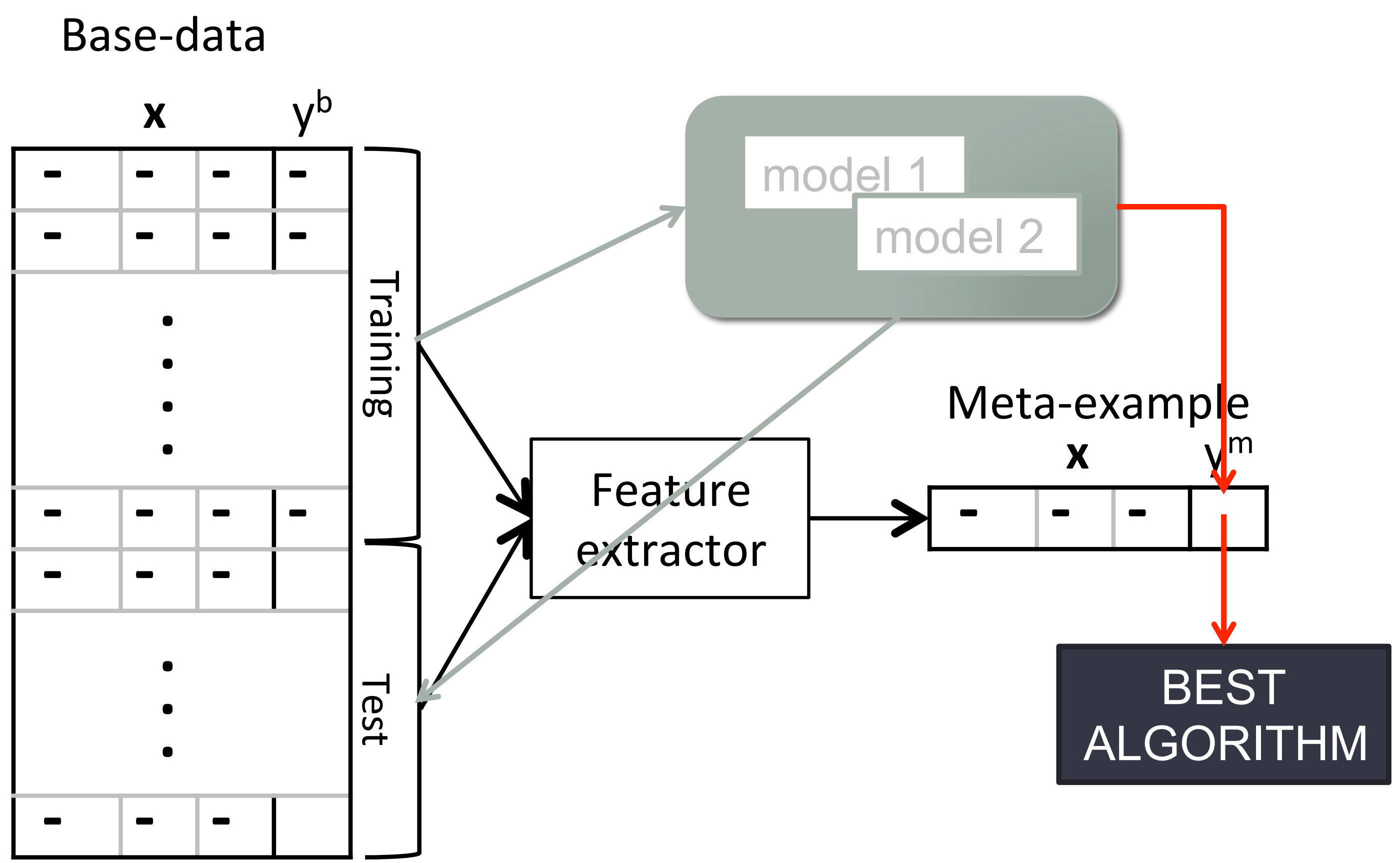
- Periodic algorithm selection for non-stationary environments
- MetaStream: a metalearning approach
  - Meta-model relates data characteristics to base-level model performance

Base-level

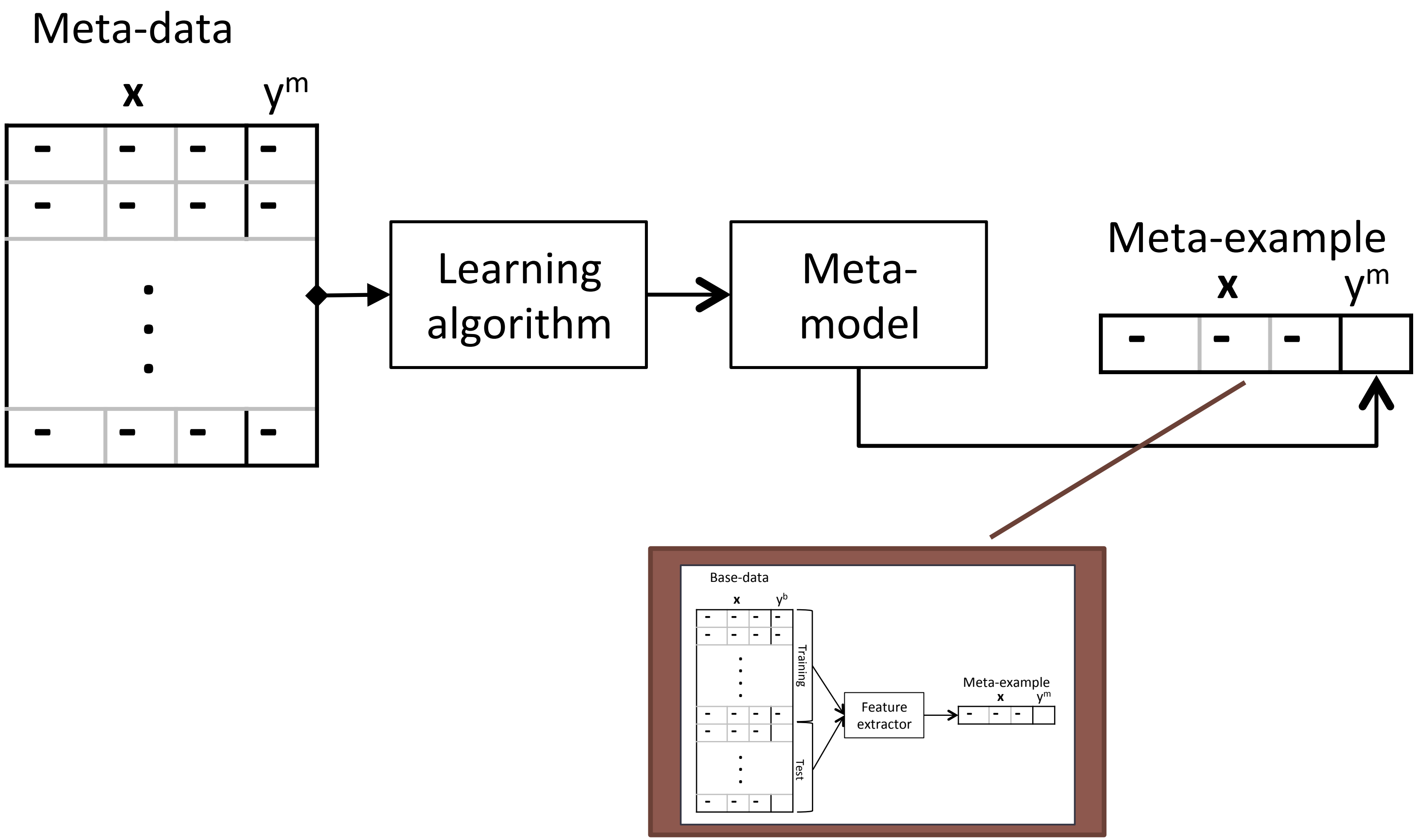


- For each new batch of *n* test examples:
  - Induces regression models using most recent training data
  - Predicts target value for test examples
  - Evaluates predictions when true target values become available

Meta-level: meta-data



Meta-level: meta-model induction and algorithm prediction



Experiments: comparison

- Meta-level
  - MetaStream
  - Default class
    - majority class (i.e., regression algorithm) in the meta training data
- Base-level
  - MetaStream and Default
    - i.e., algorithms selected by these methods
  - Ensemble: average predictions of regressors

Experimental setup: base-level

- Regression algorithms:
  - Random Forests (RF)
  - Support Vector Machines (SVM)
  - Classification and Regression Trees (CART)
  - Project Pursuit Regression (PPR)
- Data
  - Travel time prediction (TTP) problem
- Time
  - Training window: 1000 examples
  - Sliding step: 1 example
- Evaluation measure
  - NMSE

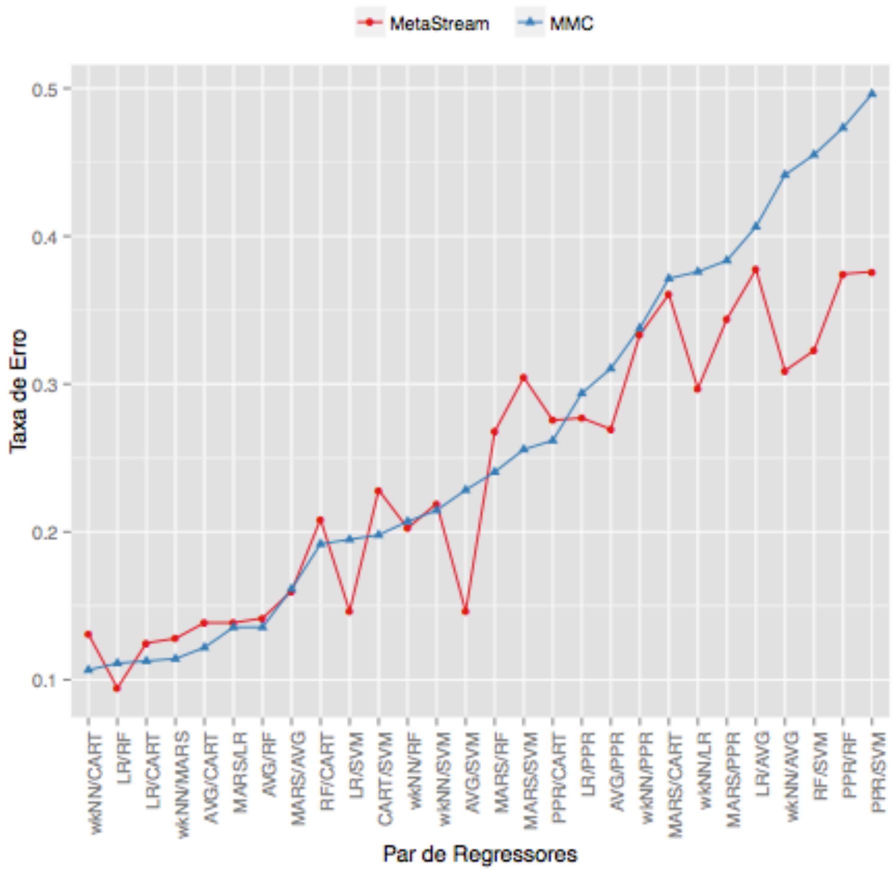
Experimental setup: meta-level data

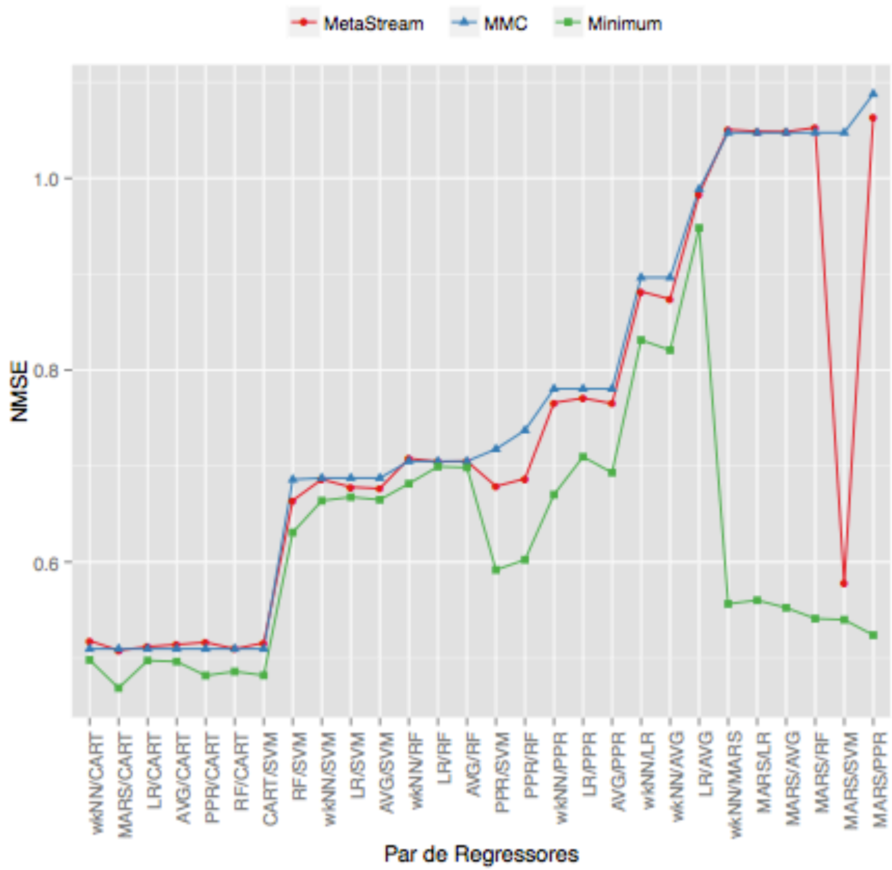
- Pairwise comparisons of algorithms
  - blocks of 25 base-level examples
  - for every pair of algorithms, which one is the best
    - or tie if difference between base-level NMSE < 0.1
  - … but no ranking
- Experiments with and without tie meta-examples
  - … on the training set
- Meta-features
  - possibility of existence of outliers, dispersion gai skewnes kurtosis, average, variance, minimum, maximum, n correlation between attribute and target
  - … for each independent variable

Experimental setup: meta-level

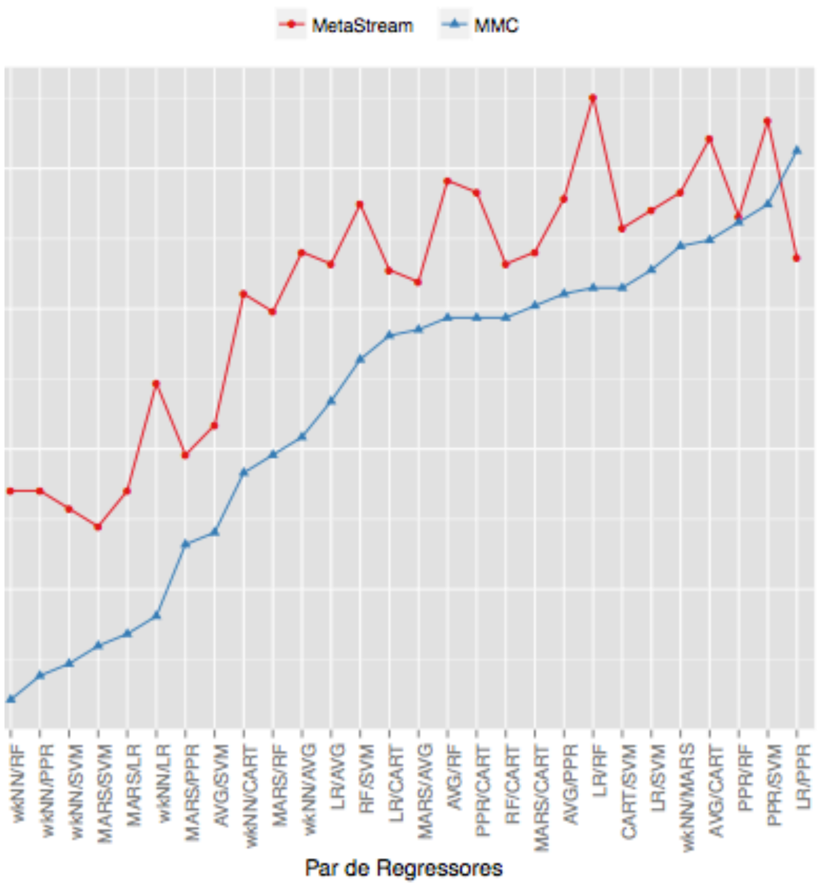
- Meta-level learning
  - classification task
  - RF
- Meta-model updating strategies:
  - Dynamic: updated for each new meta-example
  - Static: never updated
- Meta-time
  - Sliding window of 300 meta-examples
  - Sliding step of 1 meta-example

Sometimes the winner is unclear

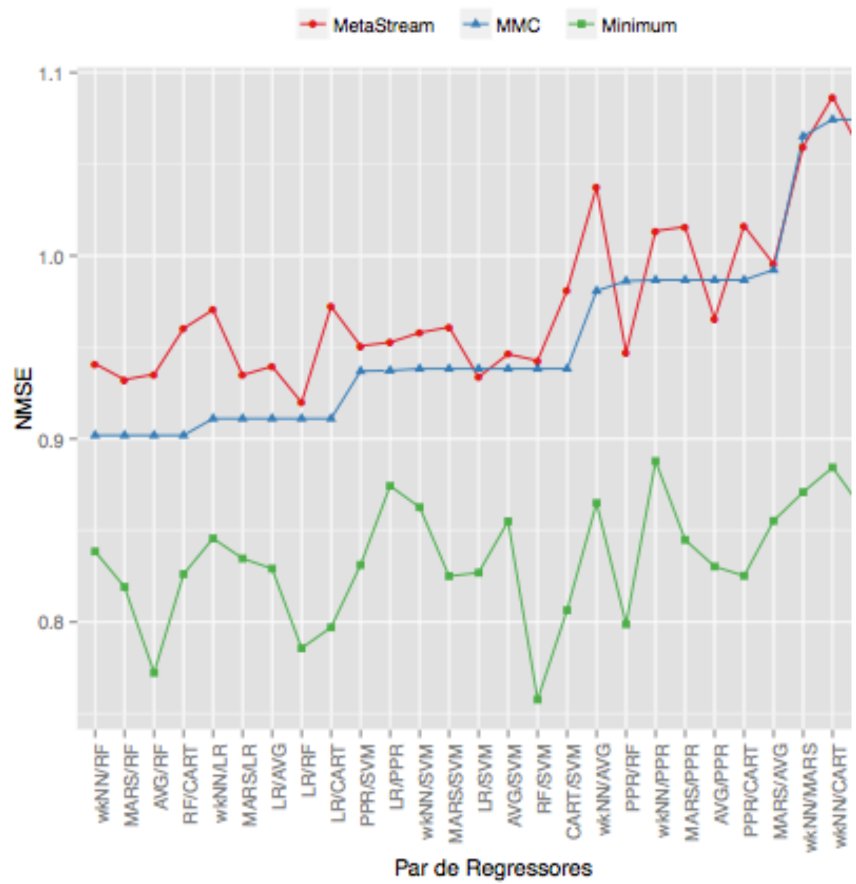




Sometimes MetaStream is worse

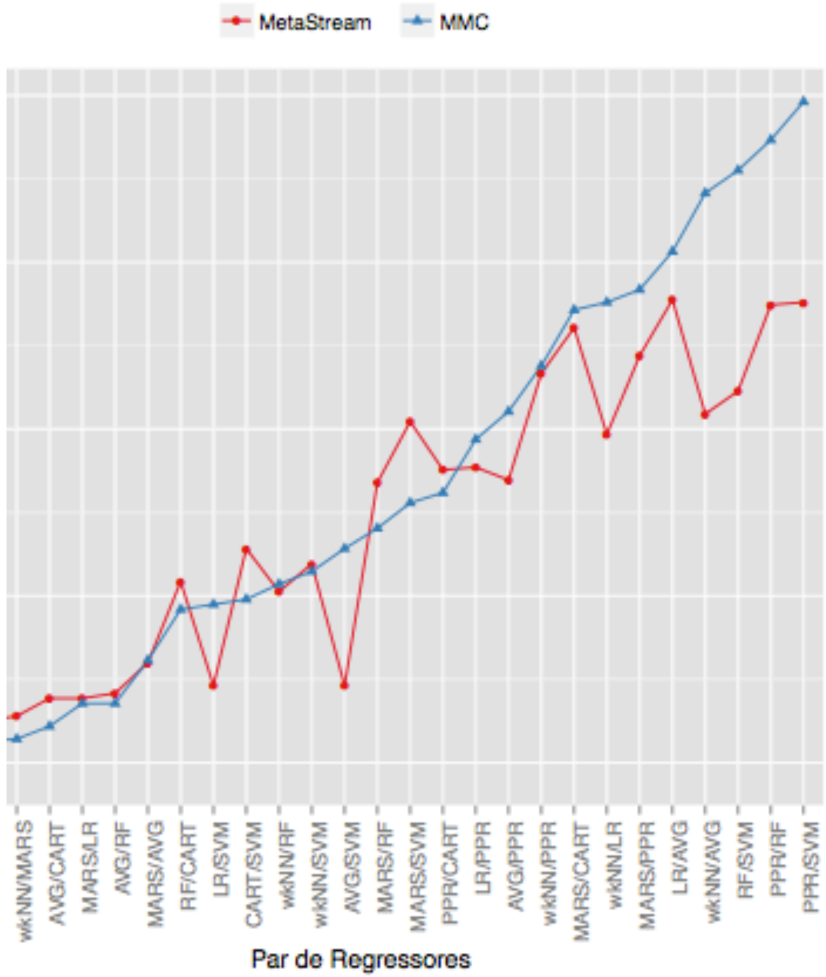


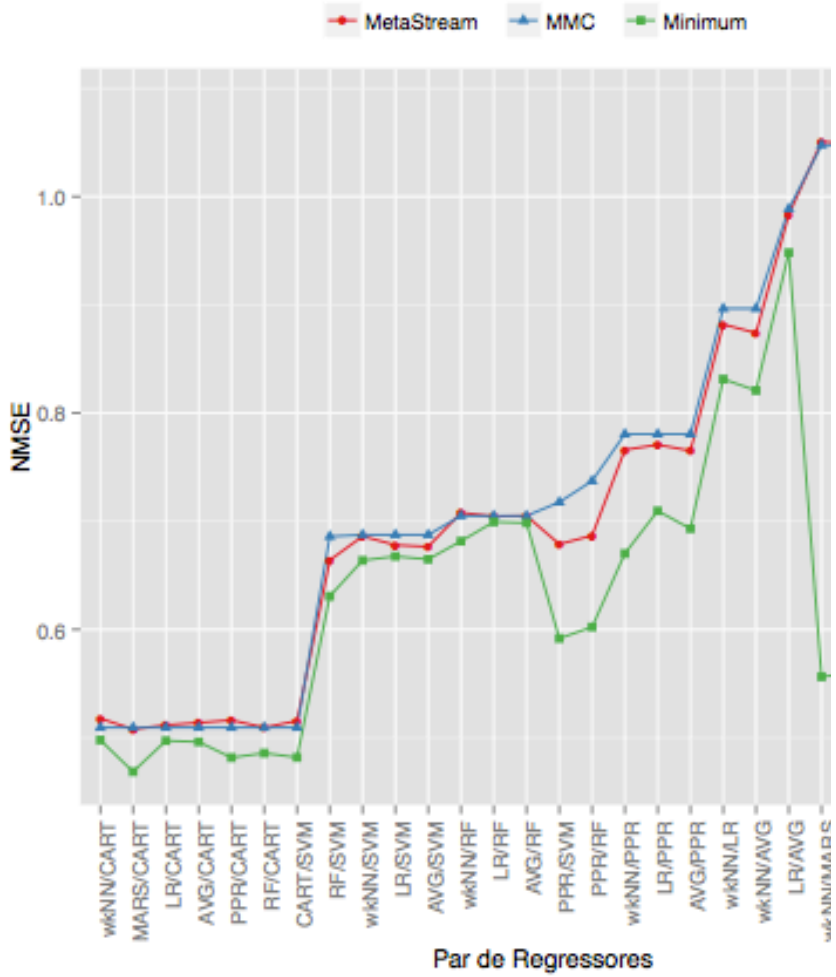
(a) airline LGA meta



(b) airline LGA base

Sometimes MetaStream is better





Meta-data issues

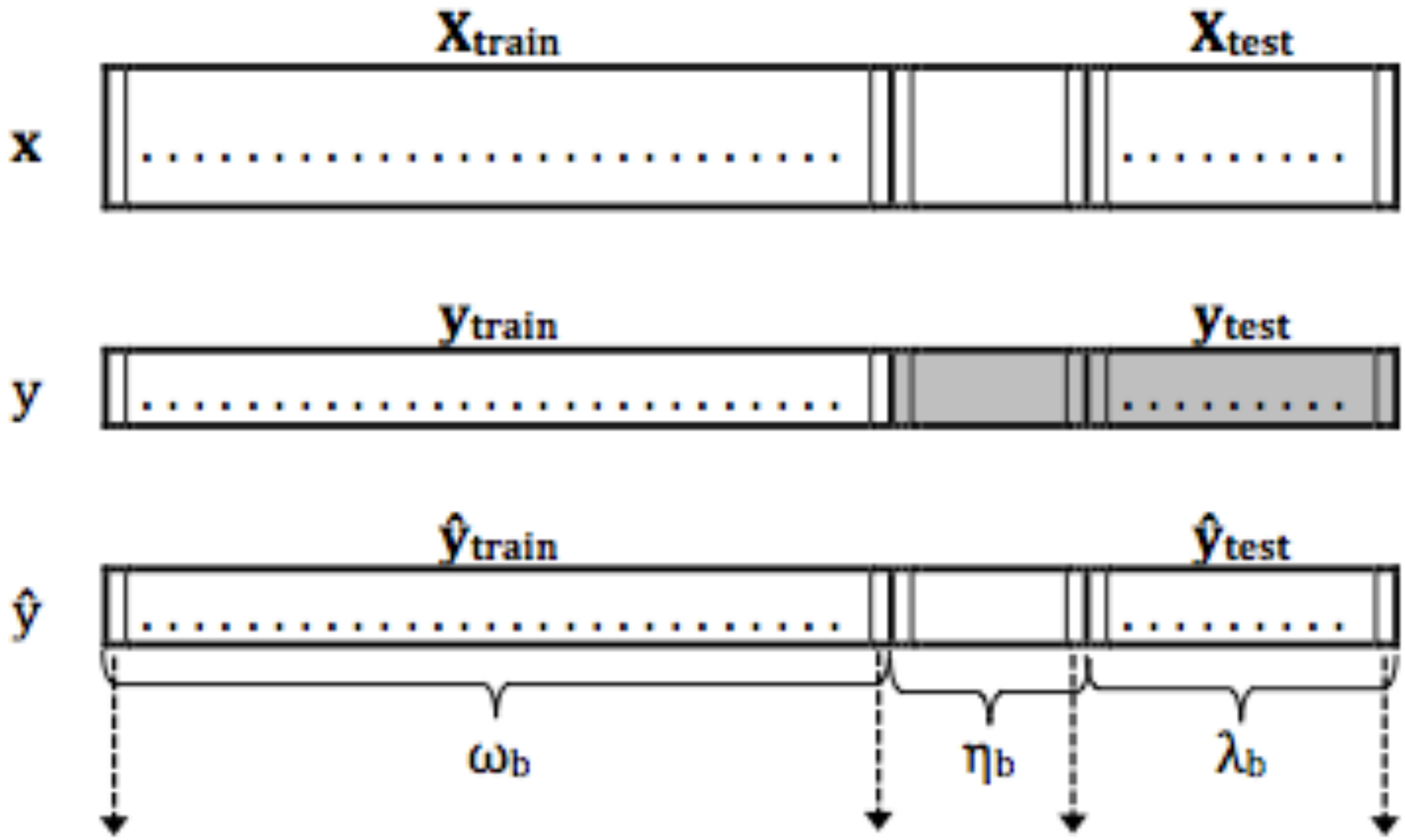
- Quality of the meta-data is sensitive to the size of the test set (δ):

Quality of data characteristics

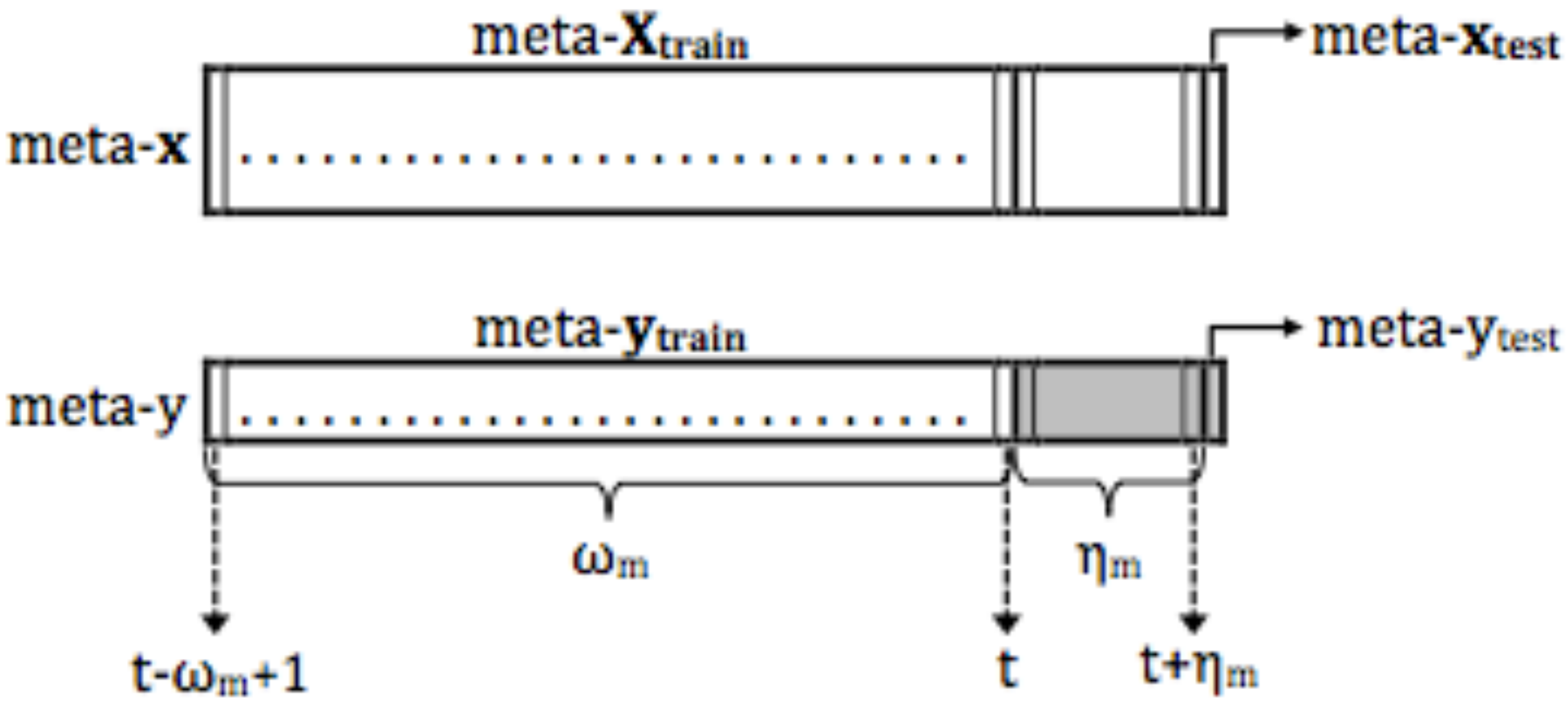
Reliability of estimation of model performance

Difference between models

Meta-features issues: base level



Meta-features issues: meta level



Conclusion and Future Work

- MetaStream is a promising approach for periodically selecting algorithms over time
- RF/CART and SVM/CART are the best choices of pairs of algorithms for the data analyzed
  - They achieved the smallest NMSE
- The proposed approach is not domain and algorithm dependent
- As future work, we plan:
  - To investigate meta-features specific for time-changing data
  - To move beyond simple pair of algorithms to multi-class classification

Acknowledgements

FAPESP

CAPES

FCT Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

References

- Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., & Gavalda, R.: New ensemble methods for evolving data streams. SIGKDD, ACM, 2009, pp. 139-148,
- Brazdil, P., Giraud-Carrier, C., Soares, C., and Vilalta, R., *Metalearning: Applications to Data Mining.* Springer Verlag, 2009
- Gama, J., Kosina, P.: Learning about the learning process. Advances in Intelligent Data Analysis. Springer-Verlag, 2011, pp. 162–172
- Klinkenberg, R.: Meta-learning, model selection, and example selection in machine learning domains with concept drift. LWA, DFKI, 2005, pp. 164–171
- Wang, H., Fan, W., Yu, P. S., & Han, J.: Mining concept-drifting data streams using ensemble classifiers. SIGKDD, 2003, pp. 226-235, ACM
- Widmer, G.: Tracking context changes through meta-learning. Machine Learning, 1997, 27(3), pp. 259-286
- Wolpert, D. H.: The lack of a priori distinctions between learning algorithms, Neural Computation, 1996, 8, pp. 1341–1390