MetaBags: Bagged Meta-Decision Trees for Regression

J. Khiari[1] L. Moreira-Matias[1] A. Shaker[1] B. Ženko[2] S. Džeroski[2]

[1]Data Science Application Group
NEC Laboratories Europe, Heidelberg, Germany

NEC

[2]Jožef Stefan Institute, Ljubljana, Slovenia

Jožef Stefan Institute

ECML/PKDD, 2018

Outline

1. Overview
   - Ensemble Learning
   - Main Challenges
2. Methodology
   - Meta-Decision Tree
   - Meta-Features
   - Bagging at the Meta-level
3. Experiments
4. Conclusion and Future Work

Plan

1. Overview
   - Ensemble Learning
   - Main Challenges
2. Methodology
   - Meta-Decision Tree
   - Meta-Features
   - Bagging at the Meta-level
3. Experiments
4. Conclusion and Future Work

Ensemble Leaning

- An **ensemble** is defined as a collection of several models that are combined to address a learning task.
- No single machine learning model can always outperform other models (so far...)
- Many winning solutions on Kaggle typically involve ensembles, not to mention the Netflix competition in 2007.

Stages of Ensemble Learning

For a given learning task, ensemble learning can be divided in three different stages:

- **Model Generation**
- **Model Pruning**
- **Model Integration**

Challenges

1. Handling regression related challenges such as the presence of outliers and an infinite target space
2. Building a model that is easily parallelizable

Plan

1. Overview
   - Ensemble Learning
   - Main Challenges
2. Methodology
   - Meta-Decision Tree
   - Meta-Features
   - Bagging at the Meta-level
3. Experiments
4. Conclusion and Future Work

Meta-Decision Tree for Regression

Definitions

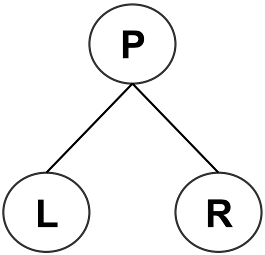
- Let the dataset $\mathbb{D}$ be defined as $(x_i, y_i) \in \mathbb{D} \subset \mathbb{R}^n \times \mathbb{R} : i = \{1, \ldots, N\}$
- Let $\hat{f}_j(x) : j = \{1, .., M\}$ be a set of M base models (*experts*) learned using one or more base learning methods over $\mathbb{D}$.
- For a given instance x and its supporting meta-features $\{z_1, \ldots, z_Q\}$, a decision tree dynamically selects the expert that should be chosen for prediction, i.e., $\hat{F}(x, z_1, \ldots, z_Q; \hat{f}_1, \ldots, \hat{f}_M) = \hat{f}_j(x)$.

Tree Induction

- We aim, at each node, at finding the feature z_j and the splitting point z_j^t that leads to the maximum reduction of impurity.
- Inductive Bias Reduction:

$$I(p) = \mathrm{IBR}(p) = \min_{j \in \{1 \ldots M\}} E\left[B\big(\mathcal{L}(p, \hat{f}_j)\big)^2\right] \tag{1}$$

where $B(\mathcal{L})$ denotes the inductive bias component of the loss $\mathcal{L}$.

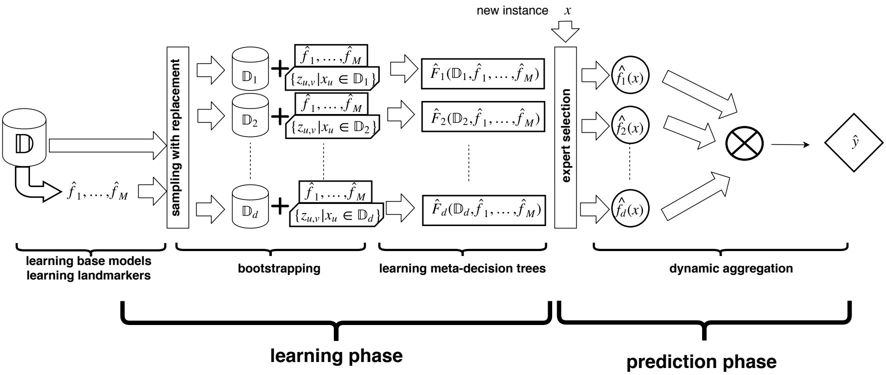


Types of Meta-Features

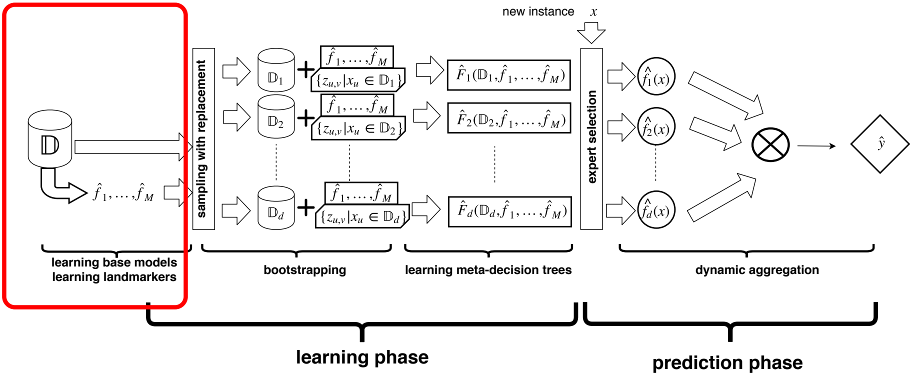
MetaBags is fed with three types of meta-features:

1. **Base features**: Following Todorovski et al, we include a dataset's base features
2. **Performance-related features**: This type of meta-features describes the performance of specific learning algorithms, e.g. the stability of predictions.
3. **Landmarking meta-features**: Derived from simple learners such as LASSO, CART, MARS, and 1-Nearest Neighbors e.g. characteristics of the landmarker in subspaces of the input.

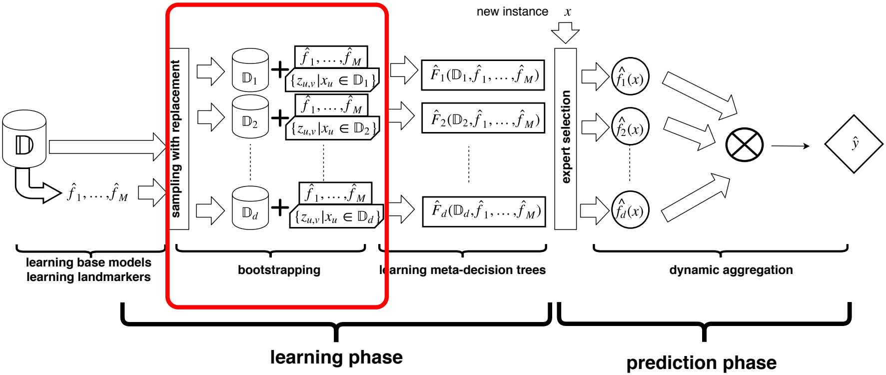
Overview of the method



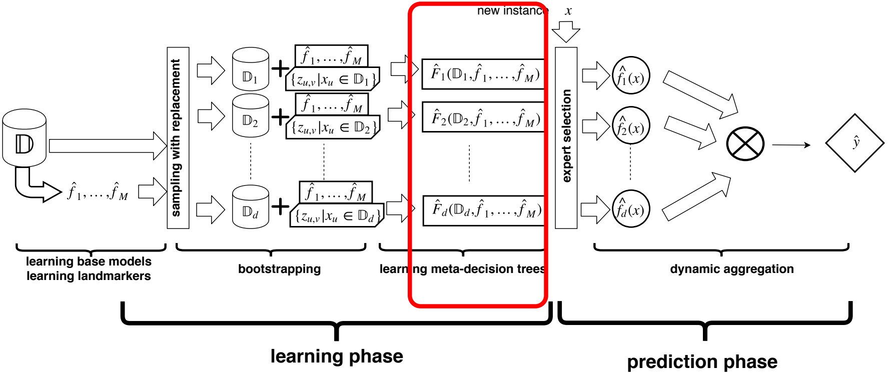
Overview of the method



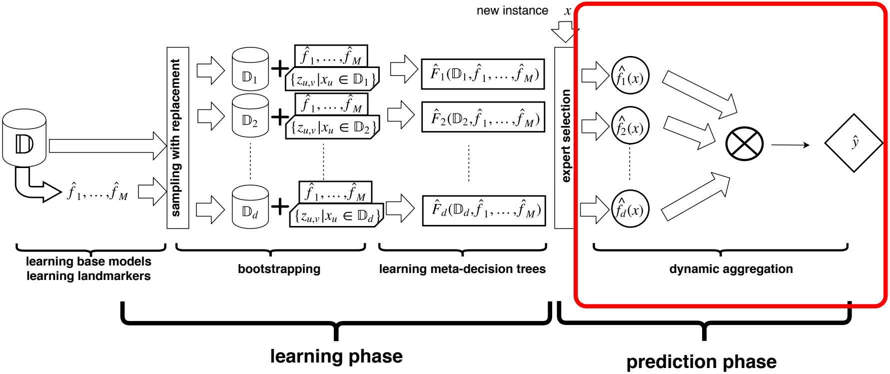
Overview of the method



Overview of the method



Overview of the method



Plan

Datasets

We have validated our methodology using:

- 4 Automated Vehicle Location datasets from Stockholm
- 13 Regression Datasets from the UCI repository

Baselines

- Base Learners: SVR, PPR, GB and RF
- Ensemble Approaches: Linear Stacking (LS), Dynamic Selection (DS), and Selecting the best individual model (Best)

Results

Dataset	SVR	PPR	RF	GB	MetaBags
R11	232.23(5.2)	242.12(8.1)	229.18(9.8)	225.45(6.5)	**220.37(4.6)**
R12	210.66(3.9)	217.34(5.4)	205.77(2.5)	200.49(6.7)	**194.36(4.2)**
R21	225.81(6.8)	240.12(5.3)	235.55(4.0)	**210.45(7.7)**	218.87(4.9)
R22	260.77(5.9)	269.19(6.5)	255.91(5.3)	**248.11(5.2)**	253.74(2.9)
C. Housing	$9.6e^{4}$($3.1e^{3}$)	$10.0e^{4}$($5.5e^{3}$)	$9.9e^{4}$($5.0e^{3}$)	**$8.9e^{4}$($3.9e^{3}$)**	$9.2e^{4}$($3.8e^{3}$)
Concrete	12.19(2.6)	15.81(2.4)	17.14(2.32)	14.54(1.7)	**11.73(0.3)**
Delta A.	$3.2e^{-4}$($1.2e^{-4}$)	$3.6e^{-4}$($2.4e^{-4}$)	$4.2e^{-4}$($2.3e^{-4}$)	$2.5e^{-4}$($1.8e^{-4}$)	**$2.2e^{-4}$($9.9e^{-5}$)**
2Dplanes	2.12(0.1)	2.57(0.2)	2.36(0.1)	**2.01(0.2)**	**2.02(0.1)**
Elevators	$6.2e^{-3}$($3.8e^{-4}$)	$6.3e^{-3}$($5.8e^{-4}$)	$6.4e^{-3}$($5.4e^{-4}$)	$6.7e^{-3}$($4.5e^{-4}$)	**$5.6e^{-3}$($5.8e^{-4}$)**
Parkinsons	$5.5e^{-2}$($7.0e^{-3}$)	$8.1e^{-2}$($7.0e^{-3}$)	$7.5e^{-2}$($6.0e^{-3}$)	$6.4e^{-2}$($6.0e^{-3}$)	**$4.9e^{-2}$($5.0e^{-3}$)**
Physic.	3.78(0.1)	3.90(0.2)	4.99(0.2)	3.70(0.2)	**3.64(0.2)**
Pole	24.11(7.2)	30.24(9.9)	28.90(5.9)	**18.54(9.3)**	20.12(3.1)
Puma32H	$3.0e^{-2}$($3.0e^{-3}$)	$2.9e^{-2}$($3.0e^{-3}$)	$2.8e^{-2}$($3.0e^{-3}$)	**$2.5e^{-2}$($6.0e^{-3}$)**	$2.7e^{-2}$($4.0e^{-3}$)
R. Wine	**0.69(0.0)**	0.91(0.0)	0.88(0.0)	0.79(0.0)	0.70(0.0)
W. Wine	0.70(0.0)	0.86(0.0)	0.76(0.0)	**0.62(0.0)**	**0.62(0.0)**
CPU_a.	**5.18(0.4)**	5.89(0.3)	6.87(0.2)	5.99(0.3)	5.45(0.2)
CPU_s.	6.07(0.1)	6.37(0.3)	8.71(0.4)	6.31(0.5)	**5.12(0.1)**
∅ Rank	2.82	4.47	4.12	2.12	**1.41**
Loss/Win	10/0	11/0	4/0	1/0	N/A

Results

Dataset	LS	DS	Best	MetaReg	MBwLM	MetaBags
R11	230.25(9.2)	222.37(5.4)	223.20(8.9)	240.67(15.7)	228.32(7.0)	**220.37(4.6)**
R12	215.32(5.9)	**190.80(3.5)**	201.46(4.3)	219.76(19.2)	211.63(7.9)	194.37(4.2)
R21	234.37(7.1)	221.98(5.8)	226.27(4.2)	249.31(6.1)	220.01(7.5)	**218.87(4.9)**
R22	260.71(4.5)	**250.37(5.5)**	254.69(5.2)	273.12(5.4)	261.41(4.3)	253.744(2.9)
C. Housing	$9.8e^{4}(4.3e^{3})$	$9.3e^{4}(3.2e^{3})$	$9.5e^{4}(4.6e^{3})$	$9.8e^{4}(4.1e^{3})$	$9.5e^{4}(3.7e^{3})$	$\mathbf{9.2e^{4}(3.7e^{3})}$
Concrete	12.14(0.3)	**11.70(0.3)**	12.00(0.3)	12.57(0.3)	11.91(0.3)	11.73(0.3)
Delta A.	$2.8e^{-4}(2.3e^{-4})$	$2.6e^{-4}(1.8e^{-4})$	$3.0e^{-4}(1.8e^{-4})$	$3.5e^{-4}(1.7e^{-4})$	$2.3e^{-4}(1.8e^{-4})$	$\mathbf{2.2e^{-4}(9.9e^{-5})}$
2Dplanes	2.35(0.2)	2.26(0.1)	2.34(0.1)	2.39(0.1)	2.15(0.1)	**2.02(0.1)**
Elevators	$6.2e^{-3}(5.5e^{-4})$	$5.5e^{-3}(7.0e^{-4})$	$7.4e^{-3}(6.1e^{-4})$	$6.5e^{-3}(4.6e^{-4})$	$5.5e^{-3}(6.8e^{-4})$	$\mathbf{5.7e^{-3}(5.8e^{-4})}$
Parkinsons	$5.9e^{-2}(6.0e^{-3})$	$6.5e^{-2}(0.0e^{-3})$	$6.2e^{-2}(5.0e^{-3})$	$5.4e^{-2}(5.0e^{-3})$	$5.1e^{-2}(5.0e^{-3})$	$\mathbf{4.9e^{-2}(5.0e^{-3})}$
Physic.	**3.46(0.2)**	3.74(0.2)	3.79(0.2)	3.80(0.2)	3.71(0.2)	3.64(0.2)
Pole	26.22(3.3)	22.15(3.6)	21.61(3.3)	25.96(4.3)	20.91(3.6)	**20.12(3.1)**
Puma32H	$4.0e^{-2}(5.0e^{-3})$	$3.2e^{-2}(6.0e^{-3})$	$3.4e^{-2}(4.0e^{-3})$	$3.3e^{-2}(6.0e^{-3})$	$2.9e^{-2}(4.0e^{-3})$	$\mathbf{2.7e^{-2}(4.0e^{-3})}$
R. Wine	0.87(0.0)	0.75(0.0)	**0.67(0.0)**	0.76(0.0)	0.70(0.0)	0.70(0.0)
W. Wine	0.82(0.0)	0.63(0.0)	0.67(0.0)	0.76(0.0)	**0.61(0.0)**	0.62(0.0)
CPU_a.	5.92(0.2)	5.93(5.5)	5.58(0.3)	5.54(0.3)	5.57(0.2)	**5.45(0.2)**
CPU_s.	5.92(0.2)	6.08(0.3)	5.90(0.2)	6.12(0.3)	5.30(0.2)	**5.12(0.2)**
∅ **Rank**	4.76	3.06	3.88	5.12	2.65	**1.47**
Loss/Win	4/0	2/0	4/0	6/0	0/0	N/A

Results

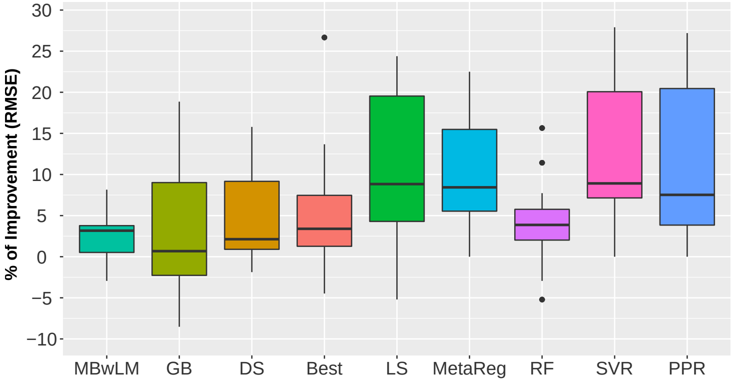


Figure: Summary results of MetaBags using the percentage of improvement over its competitors. Note the consistently positive mean over all methods.

Plan

Conclusion and Future Work

- MetaBags is a novel stacking framework for regression which uses meta-decision trees

Conclusion and Future Work

- MetaBags is a novel stacking framework for regression which uses meta-decision trees
- An exhaustive empirical evaluation, including 17 datasets and multiple comparison algorithms

Conclusion and Future Work

- MetaBags is a novel stacking framework for regression which uses meta-decision trees
- An exhaustive empirical evaluation, including 17 datasets and multiple comparison algorithms
- In our future work, we aim to adapt our method to real-time contexts and address the model generation phase.

Thank you!