Meta-Regression based Pool Size Prediction Scheme for Dynamic Selection of Classifiers

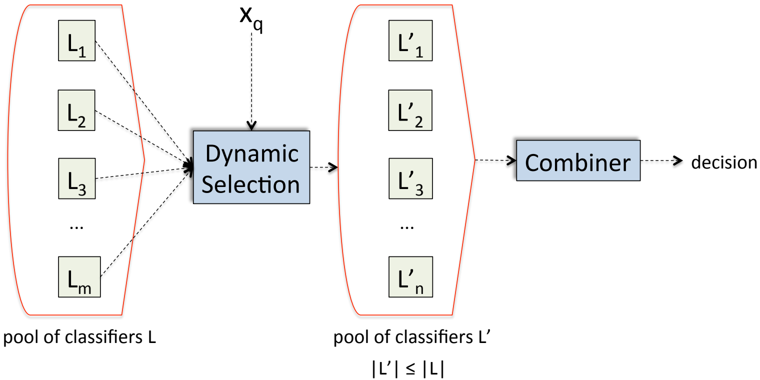
Anandarup Roy[†], Rafael M.O. Cruz[†], Robert Sabourin[†] and George D. C. Cavalcanti[⋆]

[†]École de technologie supérieure, Quebec, Canada
[⋆]Universidade Federal de Pernambuco, PE, Brazil

roy.anandarup@gmail.com (Anandarup Roy)
rafaelmenelau@gmail.com (Rafael M.O. Cruz)
Robert.Sabourin@etsmtl.ca (Robert Sabourin)
gdcc@cin.ufpe.br (George D. C. Cavalcanti)

July 29, 2016

Overview

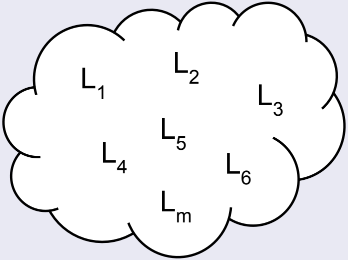


Variants: DCS and DES

- $|L'| = 1 \Rightarrow$ Dynamic Classifier Selection (DCS)
- $|L'| > 1 \Rightarrow$ Dynamic Ensemble Selection (DES)

Objective

The pool of classifiers



Size of this pool (m) is user defined yet crucial to control the computational complexity and performance of a DS.

Objective: How to find m ?

Our Approach

- A meta-regression model to **predict** pool size m.
- Our method is based on the concept of **meta-learning**.
- We emphasize **classification complexity** of a problem.
- Use regression to associate classification complexity with pool size.

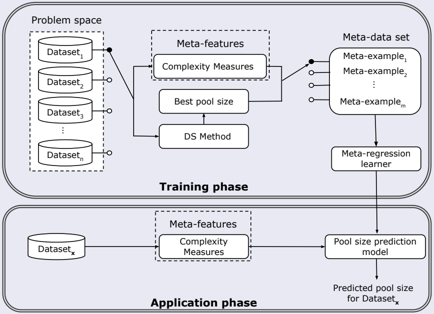
Data complexity measures

Taxonomy of the complexity measures	
Measures of overlap in feature from different classes	
$F1$	Maximum Fishers discriminant ratio
$F1v$	Directional-vector maximum Fishers discriminant ratio
$F2$	Overlap of the per-class bounding boxes
$F3$	Maximum individual feature efficiency
$F4$	Collective feature efficiency
Measures of separability of classes	
$L1$	Minimized sum of error distance of a linear classifier
$L2$	Training error of a linear classifier
$N1$	Fraction of points on the class boundary
$N2$	Ratio of average intra/inter class 1-NN distance
$N3$	Leave-one-out error rate of the 1-NN
Measures of geometry, topology and density of manifolds	
$L3$	Non-linearity of a linear classifier
$N4$	Non-linearity of the one-nearest neighbor classifier
$T1$	Fraction of maximum covering spheres

Ho and Basu, Complexity Measures of Supervised Classification Problems, IEEE PAMI, 24 (3), 2002.

Meta-Regression Framework

Framework for pool size prediction.



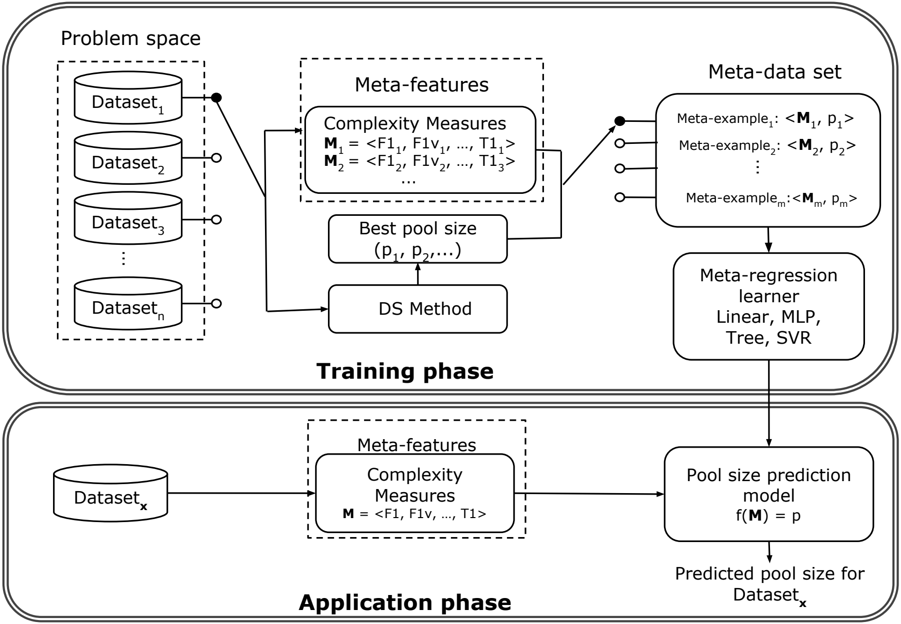
Advantage: No need to employ DS method to know pool size.

Training Phase:

- Gather large number of data sets.
- Obtain meta-features i.e., complexity measures.
- Obtain *best pool size*: the pool size with highest accuracy of DS.
- Combine to form meta-examples.
- Find a **meta-regression model**.

Application phase:

- Obtain meta-features for test data set
- **Predict** pool size using trained meta-regression model.


Problem space
Dataset1
Dataset2
Dataset3
Datasetn
Meta-features
Complexity Measures
M1 = <F11, F1v1, ..., T11>
M2 = <F12, F1v2, ..., T13>
...
Best pool size
(p1, p2,...)
DS Method
Meta-data set
Meta-example1: <M1, p1>
Meta-example2: <M2, p2>
Meta-examplem:<Mm, pm>
Meta-regression learner
Linear, MLP, Tree, SVR
Training phase
Datasetx
Meta-features
Complexity Measures
M = <F1, F1v, ..., T1>
Pool size prediction model
f(M) = p
Predicted pool size for Datasetx
Application phase

Key Features of the data sets used in the experiments

Data set	# Inst.	# Attr.	# class	Source
Adult	48842	14	2	UCI
Banana	1000	2	2	Synthetic
Blood transfusion	748	4	2	UCI
Breast (WDBC)	568	30	2	UCI
CTG	2126	21	3	UCI
Ecoli	336	7	6	UCI
German credit	1000	20	2	STATLOG
Glass	214	9	6	UCI
Haberman	306	3	2	UCI
Heart	270	13	2	STATLOG
ILPD	214	10	2	UCI
Ionosphere	315	34	2	UCI
Laryngeal1	213	16	2	LKC
Laryngeal3	353	16	3	LKC
Lithuanian	1000	2	2	Synthetic
Liver disorders	345	6	2	UCI
Magic Gamma	19020	10	2	KEEL
Mammographic	961	5	2	KEEL
Monk2	4322	6	2	KEEL
Phoneme	5404	6	2	ELENA
Pima	768	8	2	UCI
Satimage	6435	19	6	STATLOG
Sonar	208	60	2	UCI
Steel plate faults	1941	27	7	UCI
Thyroid	215	5	3	LKC
Vehicle	846	18	4	STATLOG
Vertebral column	310	6	2	UCI
WDG V1	50000	21	3	UCI
Weaning	302	17	2	LKC
Wine	178	13	3	UCI

- Repositories: UCI, STATLOG project, KEEL and Ludmila Kuncheva medical data. Two data sets are artificial.
- Multi-class data sets are divided using one-versus-all strategy to generate two-class data sets.
- Combining, we have **64** two-class data sets.
- Each data set is partitioned into 10 subsets for 10 fold CV.
- 20 pools are generated for each subset. Pools have $5, 10, 15, \ldots, 150$ perceptron classifiers.

Dynamic Classifier Selection (DCS)

- **Rank**: Classifier ranking
  - Sabourin et al., 1993
- **LCA**: Local Classifier Accuracy
  - Woods et al., 1997
- **MCB**: Multiple Classifier Behaviour
  - Giancinto, 2001

Dynamic Ensemble Selection (DES)

- **PRC**: Probabilistic Reference Classifier
  - Woloszynski 2010
- **KNORA-U**: K-Nearests Oracles-Union
  - Ko et al., Pattern Recognition 2008
- **KNORA-E**: K-Nearests Oracles-Eliminate
  - Ko et al., Pattern Recognition 2008

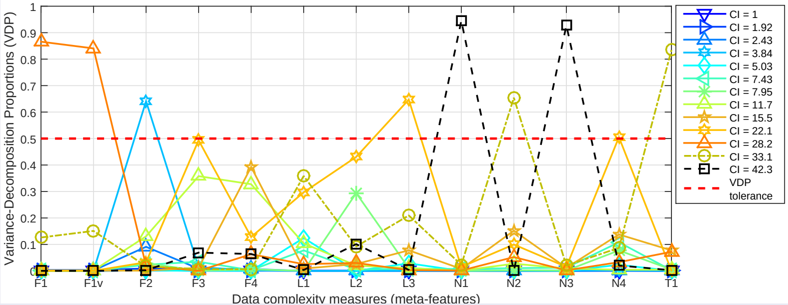
Linear and Non-Parametric Models

- Linear regression without interaction terms (Linear).
- Multivariate adaptive regression splines (MARS) without interaction (MARS-0) and with piecewise interaction (MARS-1)

Tree and Non-Linear Models

- Decision tree with binary split (Tree)
- Multi-layer perceptron (MLP) with 6 (MLP_6) and 8 (MLP_8) hidden nodes.
- Support vector regression (SVR) with linear (SVR-L) and Gaussian (SVR-G) kernels.

Belsley collinearity diagnostics



- Multicollinearity: two or more predictor variables are highly linearly related.
- The tolerances for CI and VDP are 35 and 0.5.
- A wide gap between last two CI values.
- *N*1 and *N*3 exhibit collinearity.
- Remedy: Remove *N*3 from meta-features.

Squared prediction Errors (SpE): Error between "average predicted" and the "average best" pool sizes. We perform paired t-test to compare the SpE of 64 data sets for each pair of algorithms.

Pairwise comparison between meta-regression algorithms.

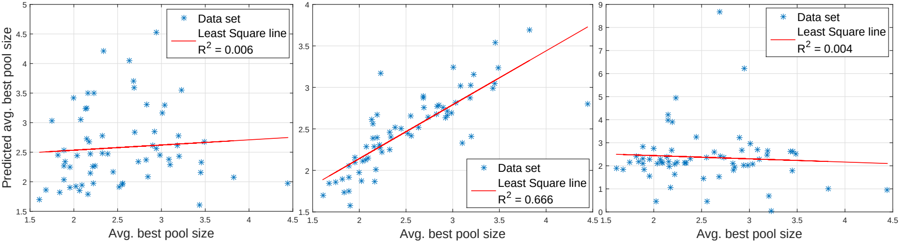
	DS methods	Meta-regression algorithms							
		Linear	Tree	MLP_6	MLP_8	SVR-L	SVR-G	MARS-0	MARS-1
DES	PRC	(1,5,1)	(0,2,5)	(5,2,0)	(4,3,0)	(1,4,2)	(0,7,0)	(0,5,2)	(1,4,2)
	KNORA-U	(1,4,2)	(1,4,2)	(6,1,0)	(6,1,0)	(1,4,2)	(0,0,7)	(1,4,2)	(1,4,2)
	KNORA-E	(1,4,2)	(0,1,6)	(5,2,0)	(6,1,0)	(1,4,2)	(0,6,1)	(1,4,2)	(1,4,2)
DCS	Rank	(0,7,0)	(0,5,2)	(2,5,0)	(3,4,0)	(0,6,1)	(0,7,0)	(0,5,2)	(0,7,0)
	LCA	(2,4,1)	(0,2,5)	(4,3,0)	(6,1,0)	(1,4,2)	(0,1,6)	(2,4,1)	(2,3,2)
	MCB	(1,6,0)	(0,4,3)	(3,4,0)	(3,4,0)	(0,7,0)	(0,7,0)	(0,5,2)	(0,5,2)

- Each cell presents amount of win, tie and loss for an algorithm over all its competitors.

Observations

- SVR-G never outperform any of its competitors.
- Tree, except one case, never outperform any of its competitors.
- Inferior prediction performance of Tree and SVR-G.
- MLP_6 and MLP_8 outperform their competitors in many cases.

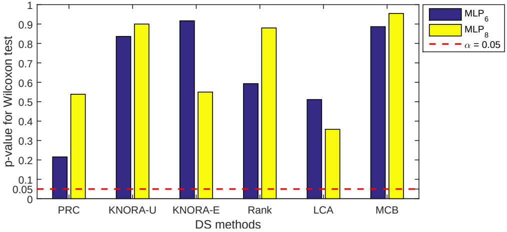
Predict the average pool size: Performance evaluation
Illustration for PRC method



Tree $\qquad$ MLP_6 $\qquad$ SVR-G

Observations

- Predictions using Tree are more scattered ($\rho = 0.08$) compare to MLP_6 ($\rho = 0.82$).
- Predictions yield by SVR-G are also scattered ($\rho = -0.07$).
- For SVR-G: Negative slope of the regression line $\Rightarrow$ unreliable predictions.
- MLP_6 provides predictions close to the actual.



Wilcoxon signed rank test to assess equivalence between classification accuracies between DS with best pool size (DS_b) and DS with predicted pool size (DS_p).

Observations

- For both MLP_6 and MLP_8, DS_p performs equivalent to DS_b.
- The accuracy in prediction is reflected in the DS performance.
- DS methods are robust enough to withstand certain prediction errors.

Conclusions

- Meta-regression (specially MLP) can be an useful tool for predicting the pool size of a DS.
- Compare to brute force, our technique is cost effective.
- Complexity measures are competent meta-features for pool size prediction.

Future Scope

- More data sets to cover whole complexity space.
- Include other categories of meta-features for multi-class problems.
- Cluster the data sets according to meta-features to find more accurate regression models.
- Ensemble of regression models for a better prediction.

THANK YOU