Ordinal model reuse and selection for a varying number of categories

Adolfo Martínez-Usó, José Hernández-Orallo, Cèsar Ferri and María José Ramírez-Quintana
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
{admarus,jorallo,cferri,mramirez}@dsic.upv.es

LMCE 2015 – 11th September 2015

Overview

- Introduction
- Hybrid model and Evaluation measure
- Case study
- Experimental setting and results
- Conclusions and Future work

Introduction

- Ordinal regression (OR)
- Number of bins vary from training to deployment
- Analyse models in a range of contexts (number of bins)
- Sentiment analysis

GOOD ⇨ BAD

GOOD ⇨ NEUTRAL ⇨ BAD

EXCELLENT ⇨ GOOD ⇨ FAIR ⇨ POOR ⇨ VERY BAD

Introduction

- Addressing the problem of considering different number of sentiment categories in a real-valued sentiment analysis approach.

- Running example for a real application problem.
  - New hybrid method that combines several regression techniques.
  - Apply this methodology to a wide range of regression datasets.

- Analyse how a regression model is reused for a range of contexts → Context plots.

Hybrid Model and Evaluation

1. Given a regression dataset, we divide it into training and validation subsets. Several models are trained.

2. Different number of bins according to the operating context are added to the validation data following an equal-width setting for the bins.

3. Models are then evaluated on the validation subset for each bin and the dominant method for each number of bins is identified.

4. In deployment time and with different data, the best model for that number of bins is applied.

chist-era

REFRAME

Evaluation by Macro-Discrepancy

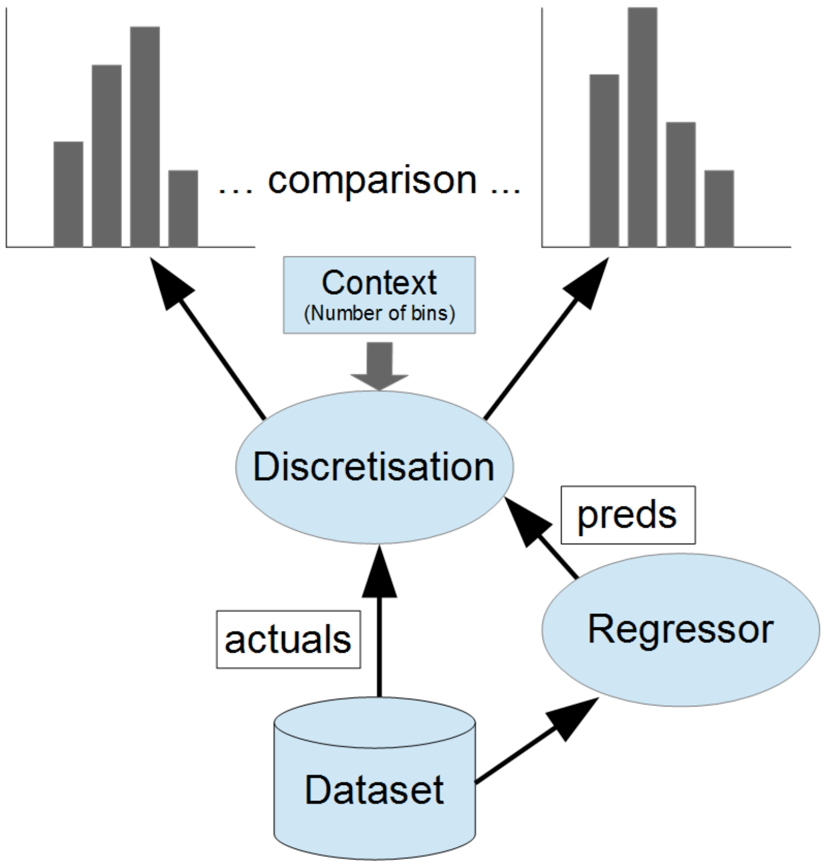
- **Measure for multiclass classification problems for which there is an inherent order between the classes.**
- **AMAE (average MAE) in the literature.**

$$mD = \frac{1}{Q}\sum_{j=1}^{Q} mD_j = \frac{1}{Q}\sum_{j=1}^{Q}\frac{1}{n_j}\sum_{i=1}^{n_j}|O(y_i) - O(\hat{y}_i)|$$

chist-era

REFRAME

Discretisation



- Adapt a regression model to several contexts.
- Each context is defined by the number of bins in which we want to classify the deployment data.

$$b_{width} = \frac{d_{max} - d_{min}}{Q}$$

chist-era

REFRAME

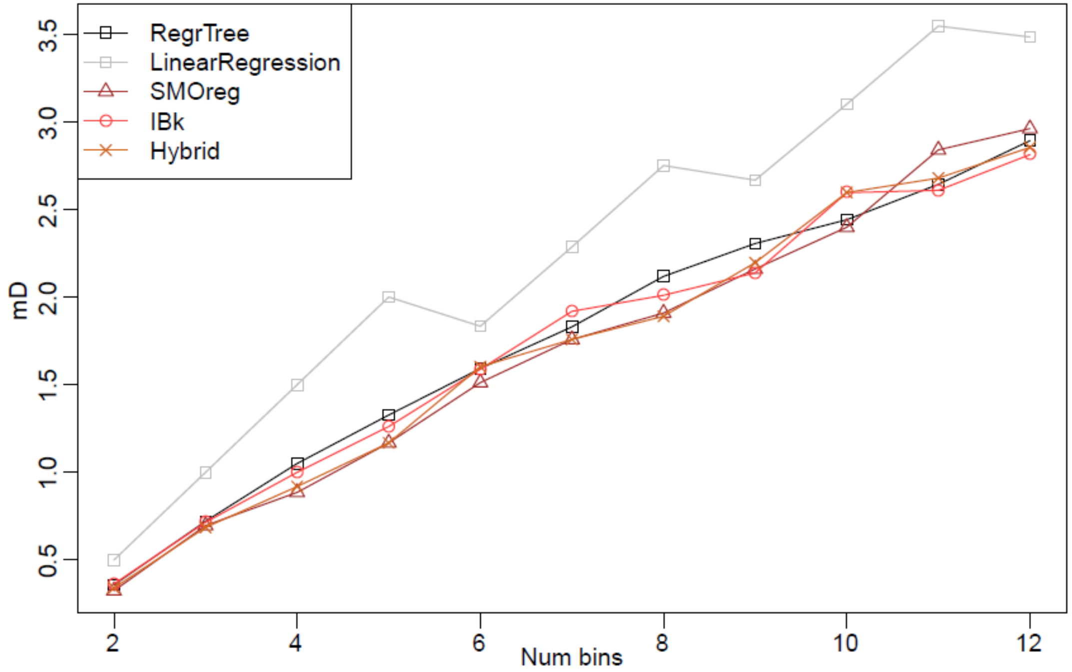
Case study

- Problem of considering different number of sentiment categories in a real-valued sentiment analysis.

- Real-world dataset of reviews of films labelled with scores on a 91 point scale (1.0;10.0;0.1).

Case study

- APC = area under the curve constructed by plotting the performance measure mD.



	RegrTree	LinearRegression	SMOreg	IBk	Hybrid
DB_SentAnalysis	17.65	22.67	**16.97**	17.42	17.09
Rank	4	5	1	3	2

chist-era

REFRAME

Experiments

Regression techniques:

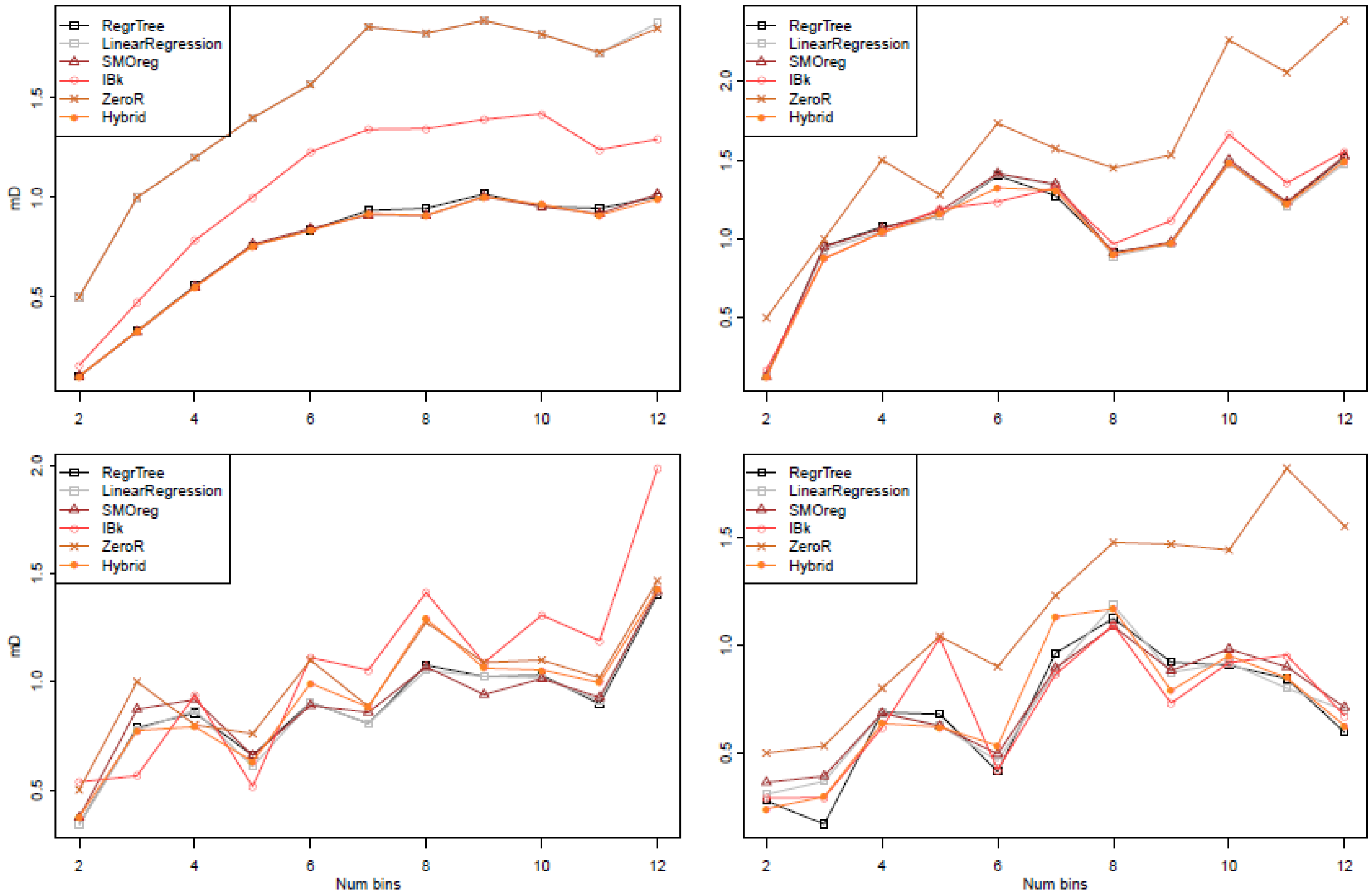
- Regression tree (RegrTree)
- Linear regression (LinearRegression)
- Support vector machine (SMOreg)
- K nearest neighbour (IBk)
- ZeroR

Dataset → 50% training, 25% validation, 25% test.

chist-era
REFRAME

Experiments

- 25 regression datasets and 5 regression techniques



Experiments

- Table shows mD values normalised by the ZeroR method.
- Hybrid approach shows a very similar performance to the best algorithm.
- Ftest stands for the average rankings of Friedman test.

DATASETS	RegrTree	LinearRegression	SMOreg	IBk	Hybrid
01airfoil_self_noise	0.5093	0.6682	0.6410	**0.4101**	0.4142
02ENB2012_data_cooling	**0.5230**	0.6151	0.6328	0.6128	0.5296
03ENB2012_data_heating	**0.2122**	0.4645	0.4895	0.4852	**0.2122**
04CCPP	0.4055	0.4713	0.4522	**0.3051**	**0.3051**
05Concrete_Data	**0.5683**	0.7470	0.7442	0.5817	0.5715
06autoMpg	0.5775	**0.5764**	0.5876	0.6131	0.5985
07housing	**0.4538**	0.5900	0.6000	0.6064	0.4959
08abalone	**0.7324**	0.7532	0.7740	0.7409	0.7349
09yacht_hydrodynamics	**0.3866**	0.6607	0.8035	0.9639	0.3904
10winequality-white	0.9093	0.9253	0.9199	**0.7830**	0.7963
11winequality-red	0.8803	0.8867	0.8931	**0.8425**	0.8616
12solar-flare_1	0.8644	**0.8570**	0.8598	0.8956	0.8655
13diabetes_numeric	0.8904	**0.8842**	0.9035	1.0433	0.9368
14dee	0.5413	**0.5326**	0.5536	0.6175	0.5533
15plastic	0.5064	1.0008	0.5008	0.7084	**0.4994**
16treasury	**0.1690**	0.2029	0.2252	0.1783	0.1817
17wankara	**0.1650**	0.2089	0.2148	0.3514	0.1817
18wizmir	**0.1045**	0.1087	0.1087	0.2574	0.1057
19cpu_small	**0.1917**	0.4185	0.3594	0.2322	**0.1917**
20auto_price	**0.6093**	0.6228	0.6377	0.6316	0.6317
21pyrim	0.7831	0.6879	0.6525	**0.4882**	0.6017
22wisconsin	**1.0126**	1.0552	1.0226	1.1940	1.0488
23delta_ailerons	**0.6932**	0.7453	0.7757	0.7314	0.7091
24delta_elevators	0.7125	0.7093	0.7211	0.7351	**0.7014**
25triazines	0.8692	0.8872	**0.8588**	0.9007	0.8643
Avg.AccumPercent.	0.5708	0.6512	0.6373	0.6364	**0.5593**
Avg.ranks (Ftest)	**1.94**	3.44	3.76	3.52	2.34

Experiments

	(1)	(2)	(3)	(4)	(5)	(6)
RegrTree (1)	-	●	●	●	●	
LinearRegression (2)	○	-			●	○
SMOreg (3)	○		-		●	○
IBk (4)	○			-	●	○
ZeroR (5)	○	○	○	○	-	○
Hybrid (6)		●	●	●	●	-

- Summary of the Wilcoxon test. •⊟ the method in the row improves the method in the column. ⊡ɸ= the method in the column improves the method in the row. Upper diagonal of level significance = 0:9, Lower diagonal level of significance = 0:95.

Conclusions and Future work

- Consider different number of sentiment categories in real-valued sentiment analysis dataset. The number of categories is only known at deployment.
- We have extended their approach with a new hybrid method and by applying this methodology to a wide range of regression datasets.
- Use of context plots + Regression to binarisation idea.
- Exploring this approach for a equal-frequency binning.
- Improving the discretisation process.

chist-era

REFRAME

Ordinal model reuse and selection for a varying number of categories

Thank you

Acknowledgements: Authors would like to thank Professor Dan Ventura for his help with the Sentiment Analysis dataset. This work was supported by the Spanish MINECO under grant TIN 2013-45732-C4-1-P, and the REFRAME project, granted by the European Coordinated Research on Long-term Challenges in Information and Communication Sciences Technologies ERA-Net (CHIST-ERA), and funded by MINECO in Spain (PCIN-2013-037) and by Generalitat Valenciana PROMETEOII2015/013.