## Categorizing the feature space for two-class imbalance learning

Rosa Sicilia, Ermanno Cordelli, and Paolo Soda Unit of Computer Systems and Bioinformatics Department of Engineering Università Campus Bio-Medico di Roma





Milan, Italy 10 | 15 January 2021

## Index



Index

Imbalanced learning Scenario

Contributions

Materials and Methods

Experimental Results

Conclusions and Future Work



Imbalance learning scenario

Contributions

Materials and Methods

Experimental Results

Conclusions and Future work

#### Learning under class skew







#### Learning under class skew





Contributions





### **Rule-based Space Characterization**



## **Rule-based Space Characterization**



#### **Meta-features**

 $0 < fr_T < 1$ 

B

- Fraction of Tomek  $(fr_T)$
- Number of same class samples  $(n_c)$
- Fraction of same class samples  $(fr_c)$

 $fr_T \rightarrow 0$ 

 $0 < fr_{c} < 1$ 

 $fr_c \rightarrow 0$ 

Η

(A/E, B, C/D, F, G, H, I, L)

(A/E, F, G, H)

F

 $n_{C} > \theta N$ 

A/E

G



**Rules** 







#### **Materials & Competitors**









**Our Proposal is** stable and effective across all datasets



	Metrics					97 ROMA
	$g = \sqrt{a}$	acc+ · acc- mean	IBA = 1	$+ \alpha \cdot (acc$	$c^+ - acc^-) \cdot acc^+ \cdot acc^-$	Index
					, , )	Imbalanced
	Iman-	Method	g	IBA	0	learning
	Davenport rank analysis	Proposal 🦞	12.58	12.54	<b>X</b>	Scenario
	Imbalanced Baselines	Imbalanced Classifier	4.82	4.3	Statistical significant	
		Bagging	7.8	6.8	difference	Contributions
		AdaBoost	9.1	8.02		
	Cost Sensitive	AdaBoostNC	9.06	7.5		Materials
		AdaC2-I	11.22	11.46	80%	and
	Boosting-based	EUSBoost	8.54	11.14		Methods
		MSMOTEBoost	11.04	10.88		
V		MSMOTEBagging	9.24	9.16	Gmean	Experimenta
	Bagging-based	OverBagging	11.96	10.52		Result
		UnderBagging	9.46	11.78		Conclusion
		llVotes	8.22 7.26 60%	60%	and	
	Ensemble	EasyEnsemble	8.32	9.48		Future Wor
		BalanceCascade	9.36	10.4		
	MEG	MES-random	3.86	3.38	IBA	COSBI
	WILS .	MES-kmeans	1.42	1.38		Computer Systems and Bioinformatics



Index

M	<i>etr</i>	irs

$$g = \sqrt{acc^+ \cdot acc^-}$$

 $IBA = 1 + \alpha \cdot (acc^+ - acc^-) \cdot acc^+ \cdot acc^-$ 



#### Index of Balanced Accuracy

Imbalanced learning Scenario

~			
	ntru	huti	onc
		JULI	UIIS
	-		

**Materials** and Methods

Experimental Results

Conclusions and **Future Work** 



	Iman-	Method	g	IBA	
	Davenport rank analysis	Proposal 🦞	12.58	12.54	0
		Imbalanced Classifier	4.82	4.3	
	Imbalanced Baselines	Bagging	7.8	6.8	Simple Bagging and
		AdaBoost	9.1	8.02	Boosting can be more
	Cost Consitivo	AdaBoostNC	9.06	7.5	effective than using a
	Cost Sensitive	AdaC2-I	11.22	11.46	imhalance
		EUSBoost	8.54	11.14	inibulance
	Boosting-based	MSMOTEBoost	11.04	10.88	
		MSMOTEBagging	9.24	9.16	
		OverBagging	11.96	10.52	
	Bagging-based	UnderBagging	9.46	11.78	
		llVotes	8.22	7.26	
		EasyEnsemble	8.32	9.48	
	Ensemble	BalanceCascade	9.36	10.4	
	MES	MES-random	3.86	3.38	
	IVIES	MES-kmeans	1.42	1.38	



Index

#### **Metrics**

$$g = \sqrt{acc^+ \cdot acc^-}$$

 $IBA = 1 + \alpha \cdot (acc^+ - acc^-) \cdot acc^+ \cdot acc^-$ 



#### Index of Balanced Accuracy

Imbalanced learning Scenario

~			
( on	trik	ודוור	nnc
COIL	ιιι	Juu	ULIS

Materials and Methods

Experimental Results

Conclusions and Future Work



	Iman-	Method	g	IBA	
	Davenport rank analysis	Proposal 🦞	12.58	12.54	0
		Imbalanced Classifier	4.82	4.3	
	Imbalanced Baselines	Bagging	7.8	6.8	Simple Bagging and
		AdaBoost	9.1	8.02	Boosting can be more
		AdaBoostNC	9.06	7.5	effective than using a
	Cost Sensitive	AdaC2-I	11.22	11.46	imbalance
	Boosting-based	EUSBoost	8.54	11.14	mbalance
		MSMOTEBoost	11.04	10.88	
		MSMOTEBagging	9.24	9.16	0
	Bagging-based	OverBagging	11.96	10.52	<b>X</b>
		UnderBagging	9.46	11.78	Proposed method beats
		llVotes	8.22	7.26	the whole category of
	Ensemble	EasyEnsemble	8.32	9.48	MES competitors
		BalanceCascade	9.36	10.4	
	MEC	MES-random	3.86	3.38	
	IVIES	MES-kmeans	1.42	1.38	







UNIVERSITA CAMPUS BIO-MEDICO DI ROMA

# Thank you for the attention