Multi-Attribute Learning With Highly Imbalanced Data

L. Viviana Beltrán Beltrán*
Mickaël Coustaty*, Nicholas Journet†, Juan C. Caicedo‡ and Antoine Doucet*
*University of La Rochelle
†University of Bordeaux
‡Broad Institute of MIT and Harvard

Poster Presentation

Overview

Motivation

Approach

- Proposed strategies
- Models

Results

Conclusions

Motivation

- Bad performance due to problems in data.
- Powerful and straightforward deep learning models not able to use classical imbalance learning strategies
- Different levels of imbalance:
  - The number of attributes per image is different.
  - The problem of unrepresented classes or ’attribute-value’ combinations.
  - Inconsistency of labels/classes pairs given by annotators which are very subjective, different, and can be incorrect.

Proposed strategies

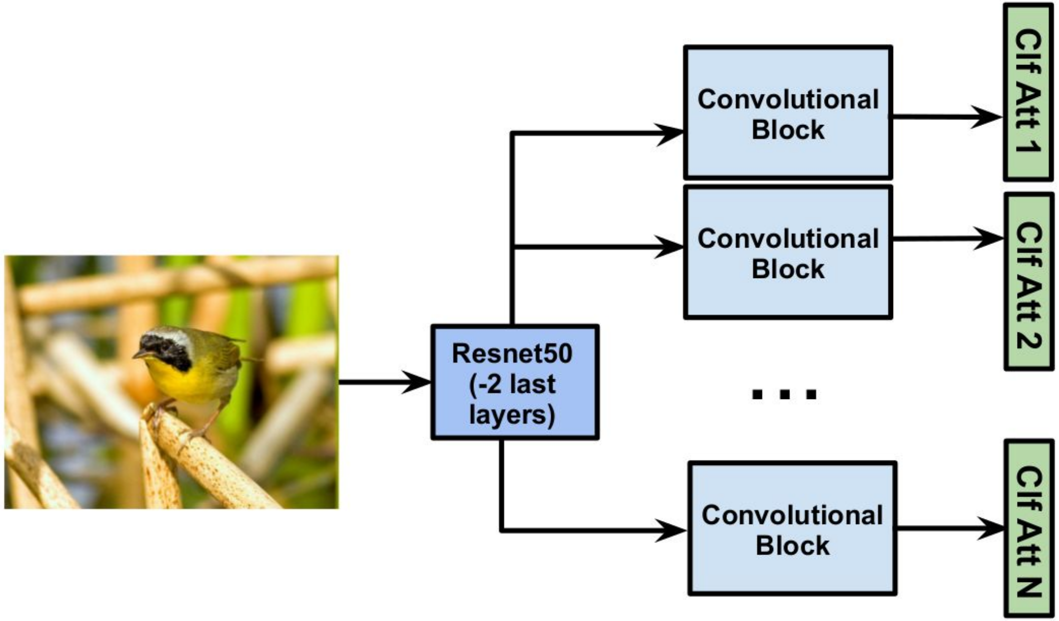
Adaptations for different levels of imbalance:

- Class-level attribute transfer: Total and partial

Adaptations for two classical imbalanced strategies:

- Sampling: Modifying the strategy to create batches → two different strategies to assign weights for samples.
- Cost-sensitive learning: Assigning weights to each class. Specialized loss function (Focal Loss)

Models



- Independent model for each attribute
- **Multitask**
- Multilabel

Results

Databases

Database	# classes	# attributes	# Train	# Test
CUB	200	312 (239 used)	5,994	5,794
AwA2	50	85	29,409	7,913
celebA	10,177	40	162,770	19,962

Incremental analysis for the performance of the 'Primary' attribute when using Sampling and cost-sensitive learning strategies.

Id	Setting	Accuracy (%)
1	None strategy	72.78
2	Data augmentation	**73.06**
3	Weighted Random Sampler with Different Weights + No Replacement	72.57
4	Weighted Random Sampler with Equal Weights + No Replacement	**73.21**
5	Weighted Random Sampler with Equal Weights + Replacement	**73.24**
6	Data augmentation + Weighted Loss	66.98
7	Weighted Random Sampler with Equal Weights + Replacement + Weighted Loss	65.70

Results for all attributes trained independently.

Attribute	Accuracy (%)
back	68.30
belly	74.49
bill	57.35
breast	72.50
crown	67.16
eye	90.69
forehead	66.41
leg	50.12
nape	66.76
primary	73.16
throat	70.50
under tail	61.46
underparts	74.51
upper tail	61.09
upper parts	69.67
wing	70.35
Average	**68.41**

Results for Multitask models.

Setting	**Accuracy (%)**
1	39.16
2	38.87
3	**39.23**
4	38.46
5	35.01
6	34.52

Best results for Multitask models - Per Attribute

Attribute	Acc (%)					
	1	**2**	**3**	**4**	**5**	**6**
back	31.77	31.44	32.17	30.96	31.20	31.08
belly	43.24	43.65	43.26	43.30	42.46	41.85
bill	35.90	35.88	35.74	35.74	35.76	35.68
breast	43.63	43.49	43.49	43.04	42.72	42.34
crown	38.68	38.60	38.79	38.02	37.98	37.31
eye	73.92	73.92	73.92	73.92	-	-
forehead	39.03	38.48	39.33	37.90	34.27	33.70
leg	22.13	21.78	21.92	21.56	19.48	18.78
nape	38.54	37.86	38.54	37.17	35.94	35.22
primary	50.58	50.08	50.67	48.67	49.33	48.41
throat	43.04	42.94	43.32	42.86	41.13	40.06
under tail	25.80	25.57	26.04	25.21	23.28	22.87
underparts	44.19	44.82	44.37	44.33	41.91	40.76
upper tail	22.17	22.05	22.55	21.96	18.31	17.73
upperparts	38.44	37.98	38.73	37.78	37.76	37.47
wing	40.02	40.04	40.14	38.60	39.49	39.65
Average	39.44	39.29	**39.56**	38.81	35.40	34.86

Comparison against state-of-the-art models in the task of multi-label.

		CUB			AwA2			CelebA		
	Model	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Macro Avg	SVM	24.15	27.52	**22.22**	58.24	59.84	51.17	50.66	**79.26**	59.50
	MLKNN	13.04	12.28	12.48	53.82	55.71	47.02	41.60	41.25	40.61
	Log Reg	29.59	10.71	14.21	58.08	60.46	51.47	65.16	46.57	51.65
	RandomF	**79.31**	1.28	1.92	64.30	51.88	46.33	58.15	23.42	27.58
	DecisionT	10.29	7.15	8.08	51.60	52.92	44.43	42.73	28.49	32.09
	ELM	7.71	**32.63**	10.34	40.38	56.68	41.06	31.52	45.52	34.73
	GaussianNB	30.30	14.54	17.04	55.32	54.08	45.95	44.32	53.98	46.53
	Deep Multilabel	23.01	17.99	18.71	**77.67**	**70.61**	**73.31**	**78.02**	70.50	**73.15**
Weighted Avg	SVM	35.49	**51.10**	41.12	72.87	66.80	67.14	67.02	**80.56**	71.42
	MLKNN	26.84	26.70	26.67	69.06	63.69	63.47	55.23	58.52	56.48
	Log Reg	47.06	21.43	26.94	73.26	66.98	67.41	72.55	62.61	65.09
	RandomF	**50.61**	7.49	9.29	74.17	63.17	62.47	65.42	45.78	49.20
	DecisionT	23.35	18.65	20.62	67.85	61.26	61.56	58.63	49.77	51.93
	ELM	21.98	37.15	25.86	61.34	56.96	57.16	52.36	56.51	53.32
	GaussianNB	41.49	26.49	30.87	69.91	62.55	62.53	57.48	64.66	59.38
	Deep Multilabel	46.00	40.73	**41.42**	**83.19**	**78.36**	**80.05**	**84.35**	77.12	**79.67**


Pink
FN
Pred
Brown
White
Yellow
Olive
FN
Pred
Olive (TP)
Gray
Yellow

Conclusions

- Study of a fine-grained attribute classification problem with CUB imbalanced database as our main use case.
- Most bad performance problems are due to the data itself.
- Straightforward deep learning models offer suitable performances.
- Adapted strategies for different levels of imbalanced.
- Adapted strategies concerning ’Sampling’ and ’Cost-Sensitive Learning approaches (same weights to all samples but with replacement).
- Loss function suitable to the imbalanced problem (Focal Loss).


Thanks for your
attention!