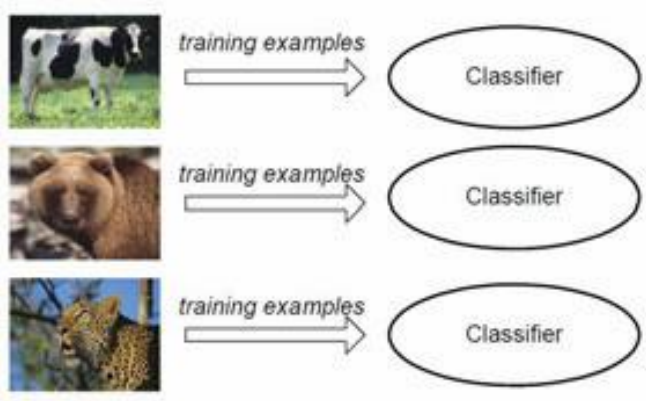
Effect of Visual Attributes on Zero Shot Learning

Jacob Chandran

Abstract

Zero shot learning is the task of classifying unknown objects by relating them to objects for which a system is familiar. In this experiment, in order to decrease error, and decrease the amount of training data necessary for a system, a system was created and tested to figure out which kind of relationships provide the best comparison between known and unknown classes, producing the least amount of error (for example, a similar visual characteristic).

Standard Image Classification



Standard Image Classification

- Uses labeled training images to classify objects
- To achieve good classification accuracy, it requires very large manually labeled training dataset
- Manually labeled dataset is costly and error prone
- When no training samples are provided for the classes, classification of unseen classes is not possible

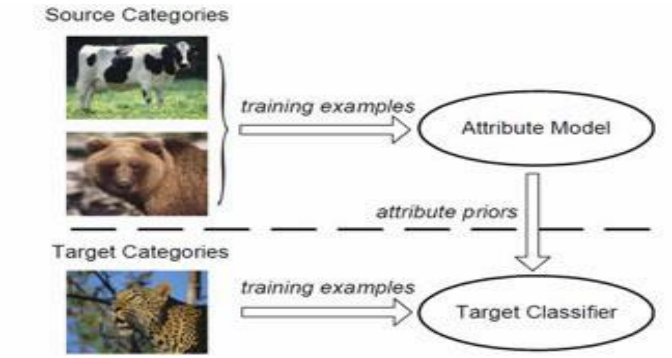
Classify about 30,000 Objects
Identify unseen class from higher level description

Large grey animal with long trunks

Human Learning

A two-wheeled self-balancing electric vehicle controlled by a stick with which you can move around while standing on top of it

Zero-Shot Learning is the ability to classify unseen visual classes when no training samples are available for the classes



Zero-Shot Learning can be accomplished using attribute classification methods

With Zero-Shot Learning, it is now possible to completely describe an unknown object which create a fairly accurate picture of the unknown classes

Is furry

Has four-legs

Mule

Has tail

Tail longer than donkeys'

ATTRIBUTES

Dog

Furry
White

Chimpanzee

Black
Big

Tiger

Striped
Yellow

Striped
Black
White
Big

Attributes are semantic descriptions
of the objects observable in images
that have human designated names .
In other words, attributes can be considered
as adjectives of the human language.

Zero-Shot Learning Applications

Useful in situations where no labeled data is available as in the case of new products, latest gadgets or new models of cars

Developing intelligent robot-learning methods.

Set of attributes that describe an object (size, weight, etc.) can be predicted from low level visual features (texture, color, etc.)

Attributes are used to infer manipulative strategies of robots: grasping objects that are small, pushing heavy objects or rolling round objects etc. thus creating intelligent robots

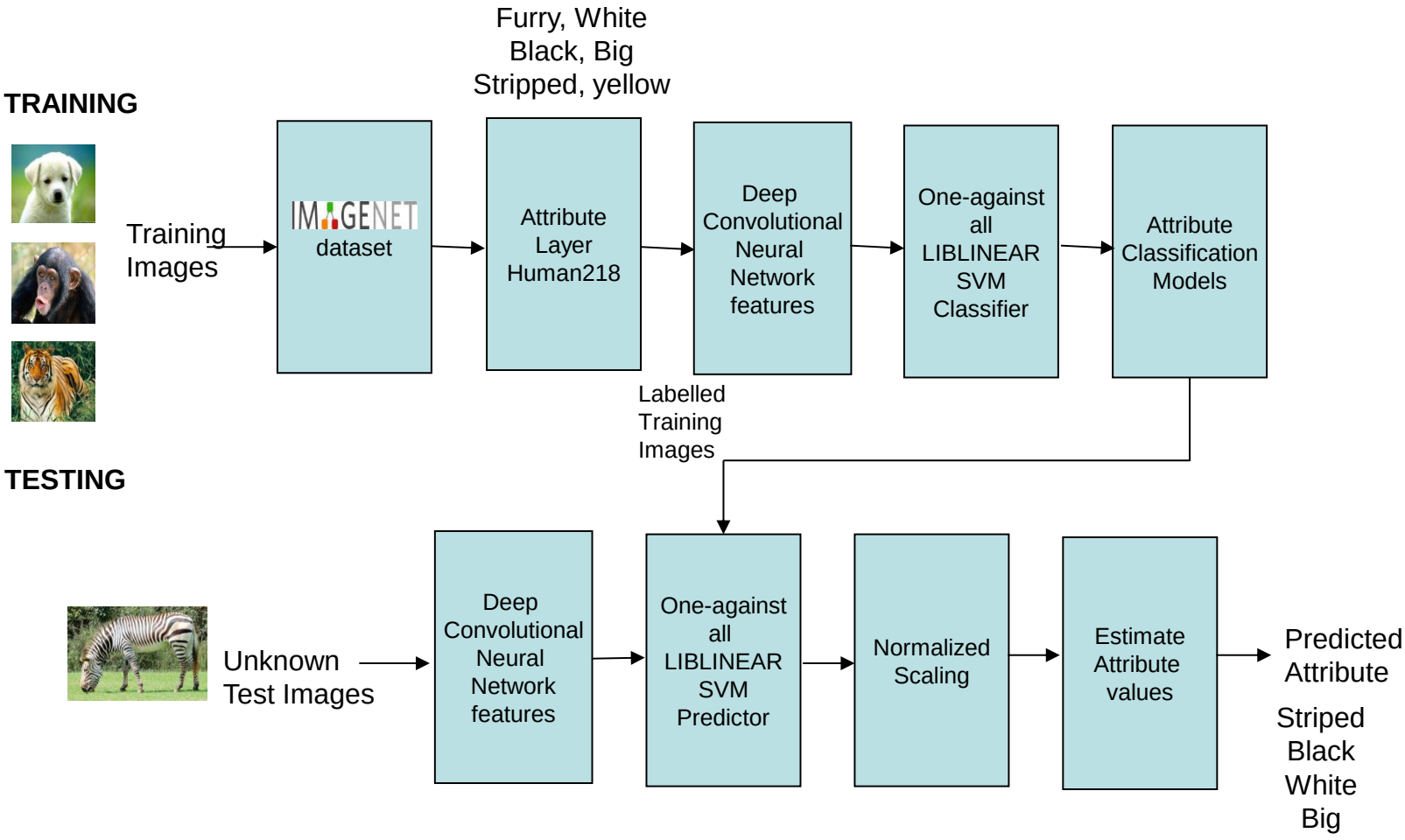
Hypothesis
Effect of visual attributes in Attribute classification accuracy

Important capability of zero short learning methods is the identification of attributes.

Certain attributes have a greater tangible and describable significance than others.

> Brightly visual attributes (e.g. Is it an animal?) contain specific and observable characteristics, should be easier to test and classify than less visual attributes (e.g. Is it man made?) which doesn't necessarily have a set of concrete definition.

The hypothesis is that visual attributes improves classification accuracy of zero shot learning systems.



Processing Pipeline of Attribute Classification Method for Zero-Shot Learning

Attribute Classification Processing Pipeline

- Training
  - Identify training image classes from ImageNet dataset
  - Build an attribute layer using Human218 attribute set
  - Compute visual features using Deep Convolutional Neural Net
  - Construct image classification models using multi-class one against all linear support vector machines from training image features.
- Testing
  - Compute visual neural network features
  - Using the model created in the training section, determine the probability that each training class is the testing class
  - Pick the training class with the highest probability, and use it to predict attributes for the test class.
  - To predict an attribute of the test class, which has no information, the attributes of a known class with the greatest similarity to the test class are used to predict the attributes of the unknown test class resulting in a zero shot classification.

Attributes, Image Data Sets

- Semantic Knowledge Base human218
  - To find nameable visual attributes for images
  - There are 218 attributes in the form of questions
  - Questions were selected to include many possible characteristics of images such as size, shape, surface properties etc. (e.g., ‘Is it man made?’ ‘Can you hold it?’).

- Semantic Attribute Mapping Matrix
  - Mapping matrix V of discrete values {-1, 0, +1} for each class. The value 1 corresponding to the class having the attribute, -1 not having the attribute and 0 for an ambiguous decision by a human annotator.
  - Class-to-attribute mapping matrix has 57 classes with 218 attribute values.

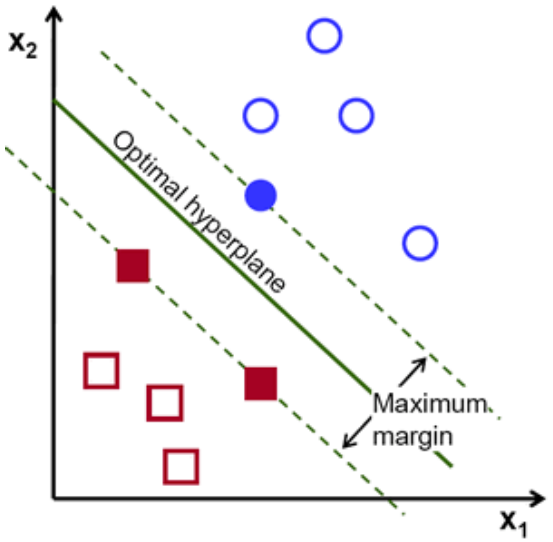
- ImageNet Image Dataset
  - Images classes built on synonyms with about 500 clean images per class
  - Uniquely linked to all nouns and has semantic relations between words for zero shot learning
  - Fifty seven image classes (e.g., ape, puppy, pizza etc) with 218 attribute questions (e.g., Can You buy it? Do you find it in a zoo?)
  - A total of 75,489 images used in zero shot training and classification

Visual Deep Neural Network Features

- Visual Deep Neural Network Features
  - Pre-Trained Deep Convolutional Neural Network Features derived from overFeat neural network architecture
  - Perform well in image classification when trained and tested with different datasets
  - Normalized 4096 features derived from RGB images of size 231 x 231

Linear Support Vector Machines

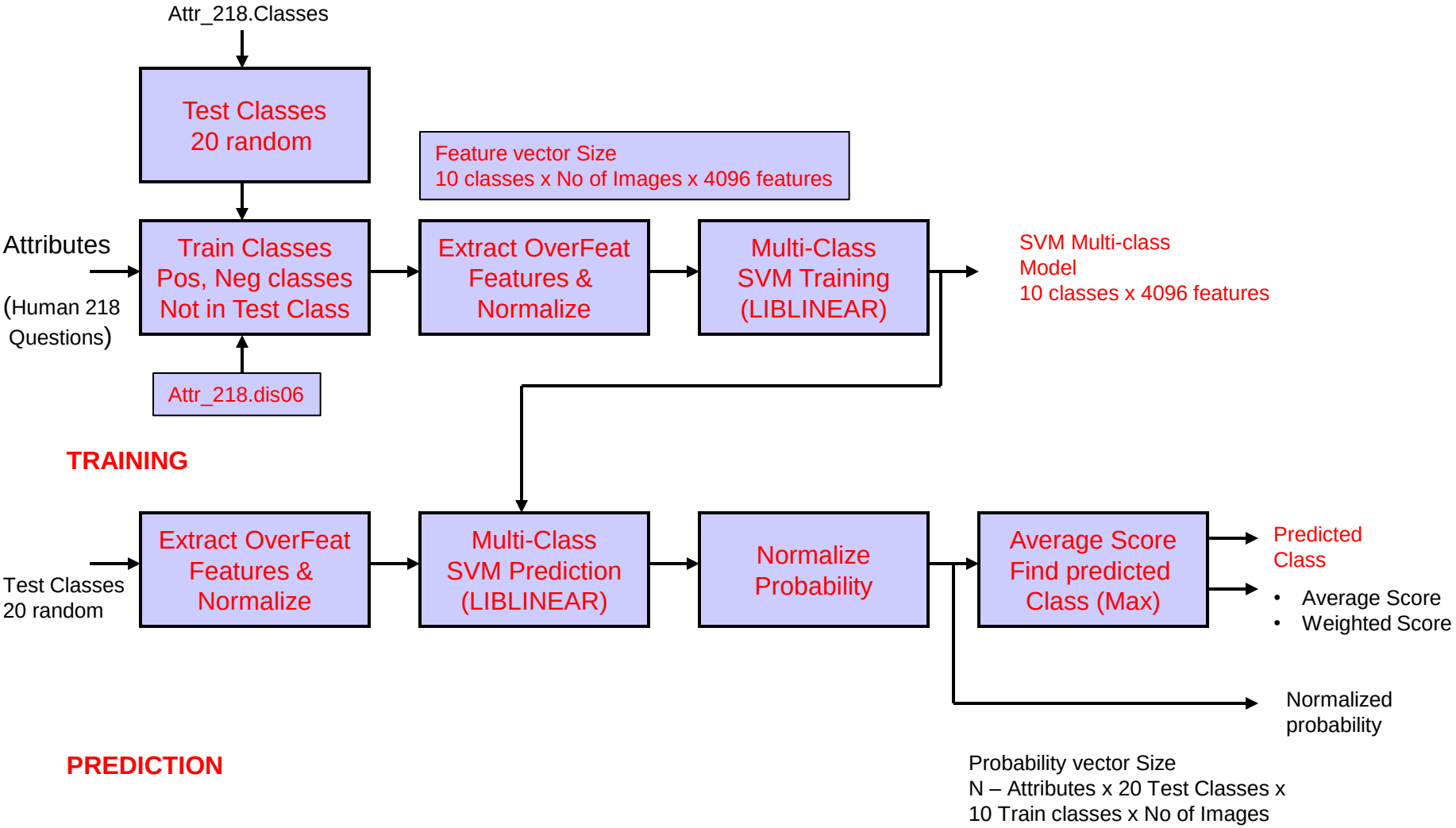
- Linear Support Vector Machines
  - Discriminative classifier defined by a separating a hyperplane
  - Given labeled training data, the algorithms outputs an optimal hyperplane which categorizes new test samples
  - Well suited to classification technique where data used for classification is large (about 57 classes, about 500 images per class with 4096 features per image)
  - LIBLINEAR, a linear SVM one-vs-all classifier is faster in training and predicting image classes



Predicting Attributes using attribute classification

- Predicting attribute values for zero shot classification
  - Convert output of one-vs-all SVM predictor to probability values using scaling
  - Probability values are used get the predicted values of the training image class labels
  - The class labels are used to estimate the attribute values {+1, 0, -1} from which the attributes are identified from the semantic knowledge base human218

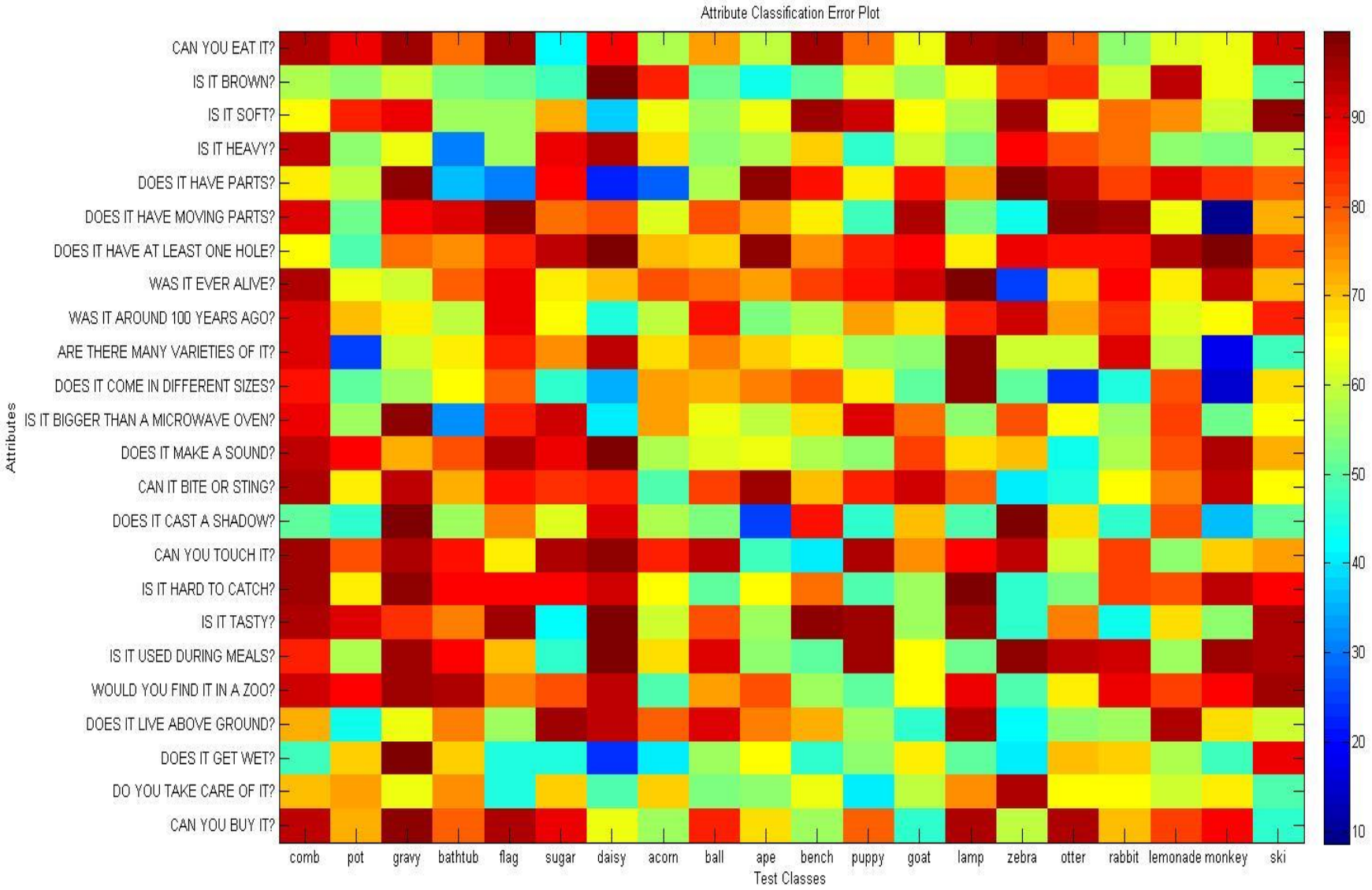
Attribute Selection Experiment



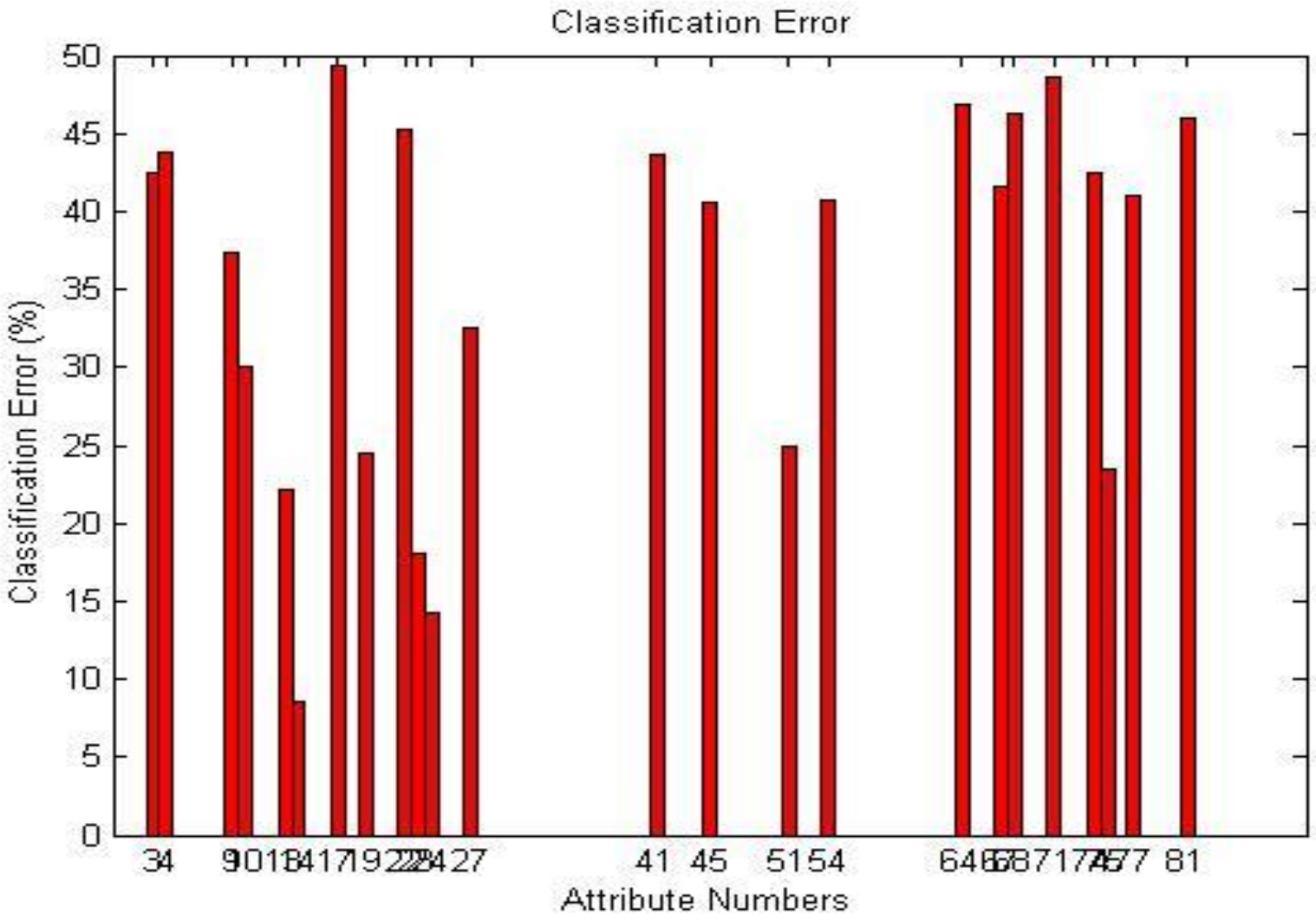
Experimental Details

- Experimental Set up
  - Choose 20 test classes randomly from the 57 image classes (e.g., lamb, zebra, comb etc.)
  - For each attribute, ten training classes not in the selected test classes were chosen.
  - The ten train classes are equally split between classes with an attribute value of +1 and -1. The value +1 indicate that the attribute satisfied visual characteristics of the image class while -1 didn't to cover the attributes in test classes
  - Create 10 Class Linear SVM Classifier Model
  - Predict the output of the attribute of the test classes and compute the error in classification

Results


Attribute Classification Error Plot
CAN YOU EAT IT?
IS IT BROWN?
IS IT SOFT?
IS IT HEAVY?
DOES IT HAVE PARTS?
DOES IT HAVE MOVING PARTS?
DOES IT HAVE AT LEAST ONE HOLE?
WAS IT EVER ALIVE?
WAS IT AROUND 100 YEARS AGO?
ARE THERE MANY VARIETIES OF IT?
DOES IT COME IN DIFFERENT SIZES?
IS IT BIGGER THAN A MICROWAVE OVEN?
DOES IT MAKE A SOUND?
CAN IT BITE OR STING?
DOES IT CAST A SHADOW?
CAN YOU TOUCH IT?
IS IT HARD TO CATCH?
IS IT TASTY?
IS IT USED DURING MEALS?
WOULD YOU FIND IT IN A ZOO?
DOES IT LIVE ABOVE GROUND?
DOES IT GET WET?
DO YOU TAKE CARE OF IT?
CAN YOU BUY IT?
Attributes
comb
pot
gravy
bathtub
flag
sugar
daisy
acorn
ball
ape
bench
puppy
goat
lamp
zebra
otter
rabbit
lemonade
monkey
ski
Test Classes
90
80
70
60
50
40
30
20
10

Results


Classification Error
Classification Error (%)
0
5
10
15
20
25
30
35
40
45
50
34
9 0 1 8 4 1 7 1 9 2 2 2 4 2 7
41 45
5 1 5 4
6 4 6 6 8 7 1 7 4 7 7 81
Attribute Numbers

Results

No	Attribute	Class	Error %
3	CAN YOU EAT IT?	sugar	42.523
4	IS IT BROWN?	ape	43.844
9	IS IT SOFT?	daisy	37.423
10	IS IT HEAVY?	bathtub	30.034
13	DOES IT HAVE PARTS?	daisy	22.209
14	DOES IT HAVE MOVING PARTS?	monkey	8.5893
17	DOES IT HAVE AT LEAST ONE HOLE?	pot	49.374
19	WAS IT EVER ALIVE?	zebra	24.559
22	WAS IT AROUND 100 YEARS AGO?	daisy	45.276
23	ARE THERE MANY VARIETIES OF IT?	monkey	17.994
24	DOES IT COME IN DIFFERENT SIZES?	monkey	14.295
27	IS IT BIGGER THAN A MICROWAVE OVEN?	bathtub	32.466
41	DOES IT MAKE A SOUND?	otter	43.697
45	CAN IT BITE OR STING?	zebra	40.638
51	DOES IT CAST A SHADOW?	ape	25.00
54	CAN YOU TOUCH IT?	bench	40.738
64	IS IT HARD TO CATCH?	zebra	46.811
67	IS IT TASTY?	sugar	41.622
68	IS IT USED DURING MEALS?	sugar	46.306
71	WOULD YOU FIND IT IN A ZOO?	zebra	48.575
74	DOES IT LIVE ABOVE GROUND?	zebra	42.469
75	DOES IT GET WET?	daisy	23.497
77	DO YOU TAKE CARE OF IT?	puppy	41.053
81	CAN YOU BUY IT	goat	46.006

Results

	Test Classes	Comb	Pot	Gravy	Bathtub	Flag	sugar	daisy	acorn	ball	ape	bench	puppy	goat	lamb	zebra	otter	rabbit	lemonade	monkey	key
No	Attribute																				
3	CAN YOU EAT IT?	95	89	97	78	97	43	87	58	74	60	96	78	63	96	98	79	54	62	63	92
4	IS IT BROWN?	57	55	61	53	52	48	99	85	53	44	51	62	56	63	82	83	61	93	64	51
9	IS IT SOFT?	64	84	88	56	56	72	37	64	56	63	96	92	65	58	96	64	77	75	60	97
10	IS IT HEAVY?	94	55	64	30	57	88	95	67	54	57	69	47	60	53	87	81	78	55	53	60
13	DOES IT HAVE PARTS?	67	59	98	36	31	88	22	28	57	98	86	66	87	72	100	95	82	91	83	79
14	DOES IT HAVE MOVING PARTS?	91	52	88	90	98	78	80	62	81	73	66	47	95	53	43	98	96	63	9	72
17	DOES IT HAVE AT LEAST ONE HOLE?	64	49	77	75	85	94	98	70	70	98	75	84	88	66	89	86	86	95	99	82
19	WAS IT EVER ALIVE?	94	64	61	79	89	67	70	80	78	73	82	86	92	99	25	69	88	66	94	71
22	WAS IT AROUND 100 YEARS AGO?	91	70	66	60	89	65	45	60	87	54	57	73	68	85	92	73	83	61	65	85
23	ARE THERE MANY VARIETIES OF IT?	91	25	61	67	85	75	94	68	76	69	67	56	55	98	60	60	91	60	18	48
24	DOES IT COME IN DIFFERENT SIZES?	86	51	56	65	79	46	35	74	72	76	80	66	50	98	51	23	46	80	14	68
27	IS IT BIGGER THAN A MICROWAVE OVEN	89	57	97	32	84	92	40	73	63	59	68	90	78	55	80	65	56	83	52	65
41	DOES IT MAKE A SOUND?	93	88	72	80	95	89	99	58	62	64	58	54	81	67	71	44	58	80	94	72
45	CAN IT BITE OR STING?	95	66	93	73	87	83	85	50	81	96	71	84	91	80	41	45	64	76	93	65
51	DOES IT CAST A SHADOW?	51	46	99	56	76	62	91	57	54	25	86	46	70	49	99	68	47	80	37	51
54	CAN YOU TOUCH IT?	96	81	95	86	66	95	97	85	93	47	41	94	74	87	93	60	81	55	69	74
64	IS IT HARD TO CATCH?	96	66	97	87	88	88	91	64	50	65	77	49	56	98	47	53	81	81	94	88
67	IS IT TASTY?	95	90	84	76	97	42	98	61	81	56	97	96	57	96	46	77	44	68	55	95
68	IS IT USED DURING MEALS?	85	57	97	88	70	46	99	68	90	54	51	96	65	53	98	94	92	57	97	94
71	WOULD YOU FIND IT IN A ZOO?	92	88	96	95	76	81	93	49	73	80	57	51	65	89	49	66	89	81	87	96
74	DOES IT LIVE ABOVE GROUND?	72	43	64	76	56	97	94	79	91	76	72	56	46	95	42	55	56	95	67	60
75	DOES IT GET WET?	48	69	99	68	44	44	23	40	57	64	46	55	66	51	41	70	69	57	48	89
77	DO YOU TAKE CARE OF IT?	71	74	64	74	45	70	49	70	53	55	64	41	60	75	94	65	65	60	66	48
81	CAN YOU BUY IT?	94	72	98	79	94	88	64	57	84	68	56	80	46	95	59	94	71	82	88	47

Conclusion

- Use indirect attribute prediction method in order to classify unknown objects and note which attributes produced the least amount of error.
- The results confirmed the hypothesis that visual attributes produce less error than non visual attributes.
- In further research, in order to decrease the amount of training data necessary to classify unknown objects, systems should be trained to identify visual attributes which will decrease the error of classification.