Sentiment Classification based on Text and Visual information

Hanqiao Lu
A20324072

Problem

Can we classify images according to sentiments based on both image itself and related comments, descriptions and titles?

Does integration of text and visual features gives additional classification ability(measured in prediction accuracy) compared with classification solely based on images or text?

Data

1. Roughly 40,000 photos are downloaded in 200 categories listed in Visual Sentiment Ontology Dataset for training ANP detectors

http://www.ee.columbia.edu/ln/dvmm/vso/download/flickr_dataset.html

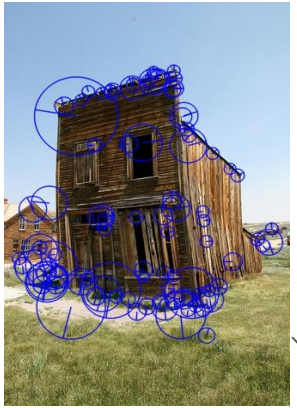
2. 500 positive and negative photos, with corresponding comments, titles and descriptions are searched and collected as training set with Flickr API
3. 200 positive and negative photos, with corresponding comments, titles and descriptions are collected as test set with Flickr API

Approach

Bag of Visual Words Processing Pipeline(BoW)

Set of photos

SURF
Keypoints
Detection



K-means
Clustering

Codebook (Visual Vocabulary)

ML
(Logistic
Regression)

Predict

FLANN based
Feature Matching

	visual words		
I1	1	2	...
...	...	...	...
Im	2	1	..
In	3	2	..

quantization

im1	w1	w2	w1	...
im2	w3	w4	w4	...
...	...	...	...	...

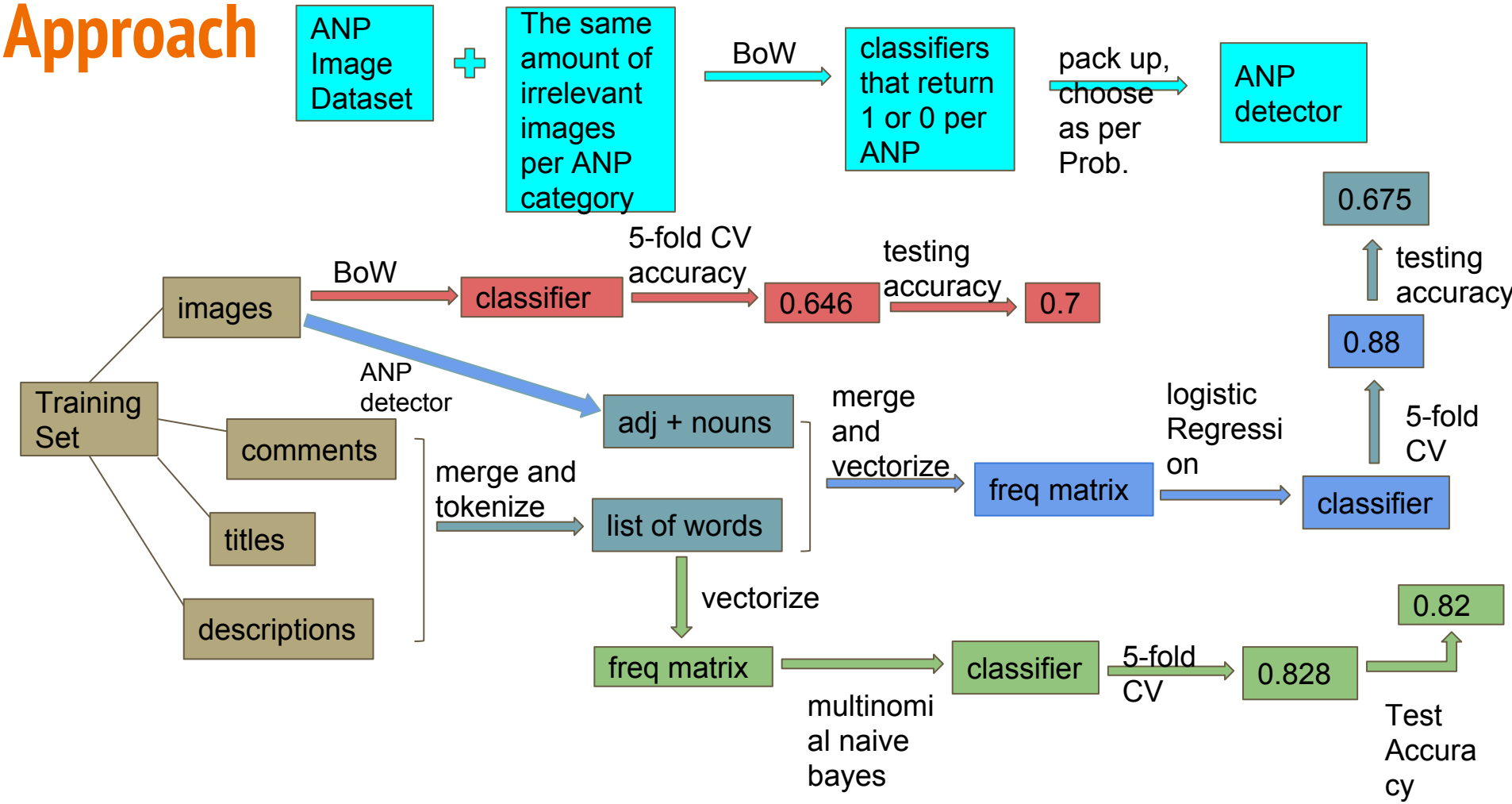
New images

SURF

FLANN

Prediction

Approach


ANP Image Dataset
The same amount of irrelevant images per ANP category
BoW
classifiers that return 1 or 0 per ANP
pack up, choose as per Prob.
ANP detector
0.675
testing accuracy
0.88
5-fold CV
logistic Regressi on
classifier
freq matrix
merge and vectorize
adj + nouns
ANP detector
images
BoW
classifier
5-fold CV accuracy
0.646
testing accuracy
0.7
Training Set
comments
titles
descriptions
merge and tokenize
list of words
vectorize
freq matrix
multinomi al naive bayes
classifier
5-fold CV
0.828
0.82
Test Accura cy

Results

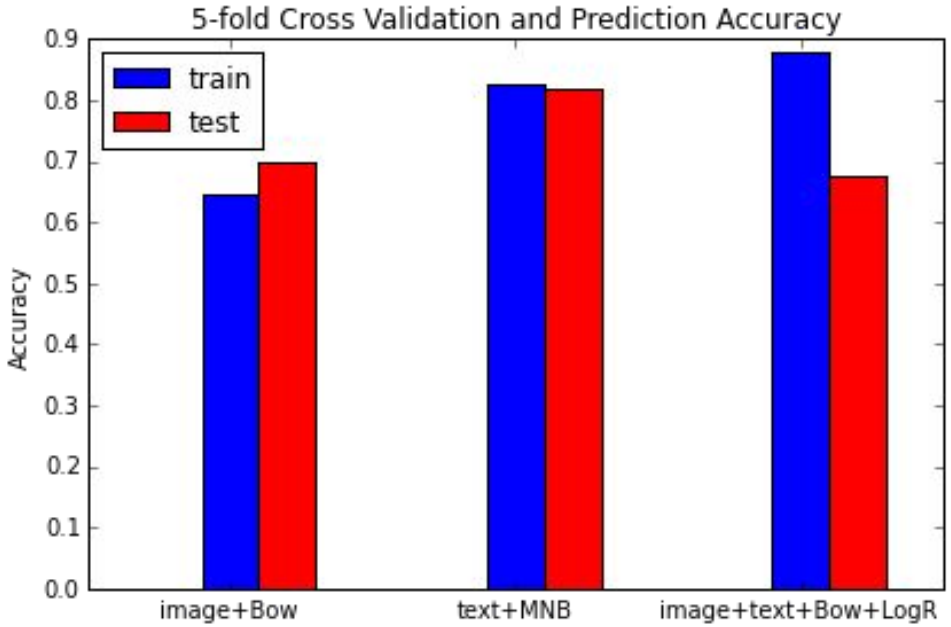


image only + BoW: 0.646, 0.7

text only + Multinomial Naive Bayes: 0.828, 0.82

image + text + BoW + logistic Regression: 0.88, 0.675

Conclusion

Data + Model	Pros	Cons
Image + BoW	Only method for photos without text information	Relatively low accuracy(0.7) Expensive computing
Text + Multinomial Naive Bayes	High accuracy (0.82) Robust (0.828 - 0.82)	Only for photos with text information(title, description or comments) Does not consider content of image(relied on people's obervation)
Text + Image + ANP detector(trained by BoW) + Logistic Regression	Can apply on all data scraped from Flickr Both image and text information are analyzed Fantastic learning capability(5-fold CV acc. 0.88)	Overfitting(0.88 - 0.675) Expensive computing High requirement of quality of training data(large, representative)