Project Report

- Sentimental Analysis on Amazon Review(Music Instruments)

By:

- Ajay Ramesh
- MSCS | IIT Chicago
- CS595

Reference

- Active Learning with Rationales for Text Classification
  - Manali Sharma, Di Zhuang and Mustafa Bilgic
  - http://www.cs.iit.edu/~ml/pdfs/sharma-naaclhlt15.pdf (http://www.cs.iit.edu/~ml/pdfs/sharma-naaclhlt15.pdf)

Date - 12/3/2017

dataset - http://jmcauley.ucsd.edu/data/amazon/ (http://jmcauley.ucsd.edu/data/amazon/)

Musical Instruments Prodcut Review is used.

Experiment Setup Overview

- Derived a small Dataset from 10 Reviews which has all kind of reviews(5, 4, 3, 2, 1 stars)
  - Train Dataset has 527 docs
  - Test Dataset has 163 docs
  - Created a seperate training and test dataset. Test is not used in any training.
- Human Labelled from My friends **Hemanth and Bashyeam.**. Around 50 Documents are carefully labelled.
- Everything is exported into csv. That is used in this experiment. Please find the data folder for datasets.

Sections

Section 1: Dataset Description

I used the Amazon review dataset on Musical Instuments, which is available in http://jmcauley.ucsd.edu/data/amazon/ (http://jmcauley.ucsd.edu/data/amazon/). The small version is present in data folder. It has below columns with human labelled rationales

Section 2: Model Fit Basic

I want to see how it behaves without active learning, so I tested across **Logistic Regression and BNB**. I choose Logistic Regression for active learning part

BernoulliNB

- Accuarcy - 81.939163498098855 %
- AUC - 100 %

Logistic Regression

C = 0.001

- Accuarcy - 65.20 %
- AUC - 61 %

C = 1

- Accuarcy - 99.61 %
- AUC - 100 %

Top 10 Feature absolute weights using Logistic Regression:

All words does not make sense The words which doesnt not make sense to me are-

- by
- almost
- midi
- thick

Its absolute weights and features are below

Out[19]:

	0	1
0	midi	1.189383
1	fine	1.072322
2	noticed	1.045061
3	came	0.970896
4	thick	0.925214
5	just	0.923685
6	almost	0.870380
7	equipment	0.860073
8	noisy	0.859717
9	by	0.837260

Section 3: Artificial Labeling

Artificial Labeling using Chi square is used

Top 10 Positive Words

Out[24]:

	0
32	amazon
683	pop
365	good
694	pretty
748	reviews
278	especially
690	practice
206	cutter
1006	with
934	try

Top 10 negative Words

In [25]:

Out[25]:

	0
961	useless
545	midi
190	control
540	mic
558	monster
543	mics
588	noisy
984	week
618	opinion
724	read

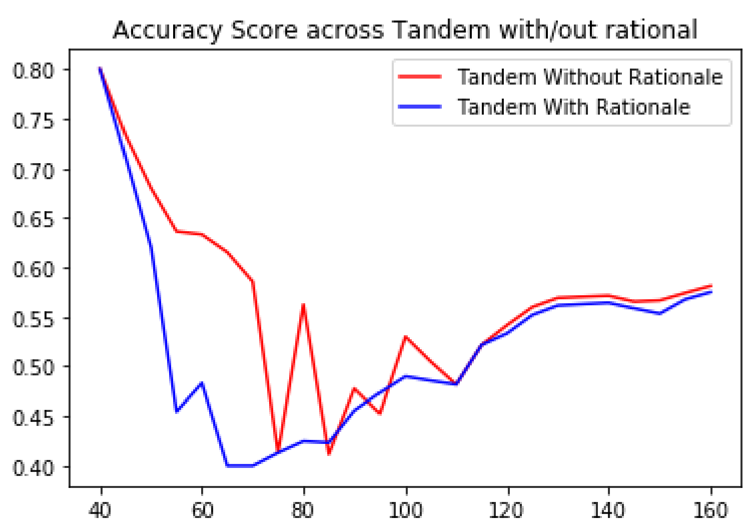
Section 4: Active Learning using Tandem Learning

Every document which has top positive word and negative word is multiplied with **r**

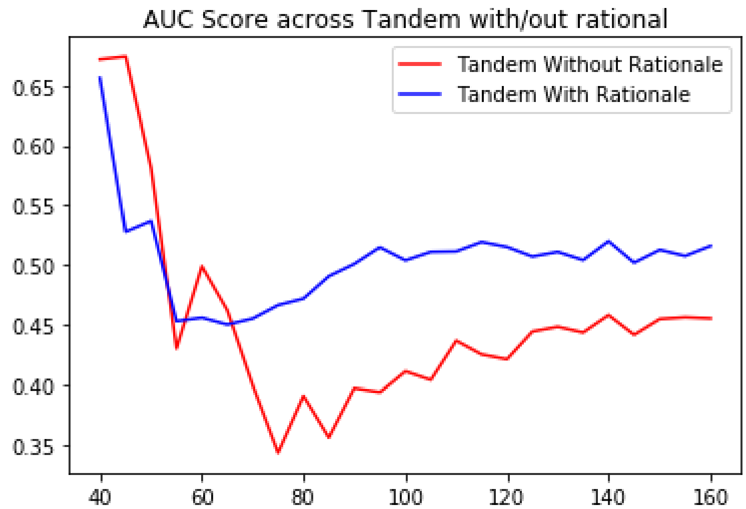
In [26]:

```
r = 1
o = 0.01
```

Out[43]:



Out[44]:



Section 5: Human Labeled Doc Preview

- Said to lable at max 3-7 words.
- on average they marked 5 words

"Not much to write about here, but it does **exactly** what it's supposed to. filters out the pop sounds. now my recordings are much more **crisp**. it is one of the **lowest** prices pop filters on amazon so might as well buy it, they honestly work the same despite their pricing,"

In [53]:

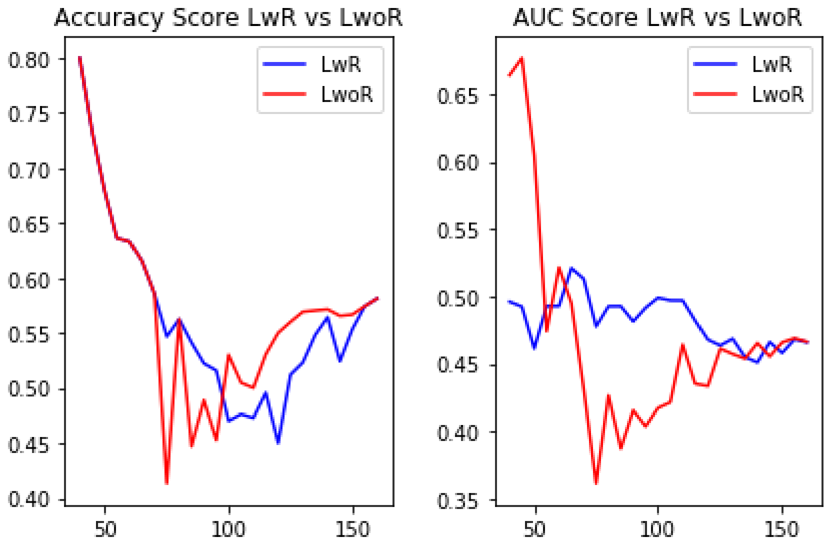
```
1 L_docs.head()
```

Out[53]:

	ReviewText	Rationales	Y
0	Not much to write about here, but it does exac...	exactly,crisp,lowest,,,,,	False
1	The product does exactly as it should and is q...	does,affordable,,,,,,	False
2	The primary job of this device is to block the...	no,noticeable,reduction,secure,,,,	False
3	Nice windscreen protects my MXL mic and preven...	prevents,pops,,,,,,	False
4	This pop filter is great. It looks and perform...	great,looks,performs,eliminate,pops,,,	False

Section 6: Active Learning with Rationale

Out[74]:



Section 7: Conclusion: Transparency Overview

LwR has great transparency below tabel sumarizes it

In [76]:

```
1 FW
```

Out[76]:

	Tandem (rational, artificial labeler)	Weights	Tandem(no rational)	Weights	LwR(Human Labelled)	Weights	LwoR	Weights
0	and	0.035263	on	0.011774	great	0.002962	on	0.011758
1	by	0.025491	to	0.008361	good	0.002933	to	0.008355
2	with	0.025165	get	0.008341	quality	0.001477	get	0.008337
3	opinion	0.018606	have	0.008263	well	0.001457	have	0.008261
4	its	0.015229	can	0.008082	cables	0.001030	in	0.007907
5	good	0.014789	in	0.007987	higher	0.001015	for	0.007708
6	control	0.010322	for	0.007722	horrible	0.001010	and	0.007377
7	has	0.008529	and	0.007383	doesnt	0.001005	can	0.006902
8	there	0.008044	other	0.006749	looks	0.001004	other	0.006714
9	sm	0.007816	little	0.006409	recommended	0.000995	little	0.006414

Section 8: Future works

- Logistic Regression is choosed just for simplicity. But in future try other models.
- Scale the dataset. Here I Used a small version of dataset
- Collect Rationales more using Uncertainity Sampling
- Extend the Web-APP UI to hold deletion.
- Try Learning with Explanation

Section 9: CD - ByProduct of IML

- CD stands for Create Dataset.
- Do we understand **fdsajfiowrjl jweiorj lwekrjoweire** text?
    - I don’t because I don’t know what language it is, also I don’t know what is the grammar to understand it. Likewise, Machine learning can’t understand random representation. So I see the bag of words model is not great for text classification. So we need to represent a good way such that ML can understand.
- The transparency achieved by LwR is great by Human Labelling part.
- I created simple flask UI which aims to create dataset for text classification for IML. Check the demo.mov on how to label it using that UI.

Section 10: Tech Stack

- Web is developed using Python Flask and Html.
- Jupyter notebook with sklearn

Bottomline

My primary motivation is understand the text, which will solve many problems today. After going through the text classification studies I see we can’t use Machine learning(For Risk Domains) with bag-of-words model which is based on pure statistics, that does not mean we have to trash Machine Learning, instead we need to re-define how we represent the text.

For text classification, I think we need to enhance how we represent text. How about an Natural Language Processing API which gives context or group of words which are important to it. No doubt LwR(learning with rationale) and LwE(learning with Explanations) will work great if the underlying representation of text is meaningful.

More labels from Human than artifical will be transparent and trustworthy.