Detecting Valence from Twitter Data

SemEval-2018 Task 1: Affect in Tweets (Sub task 4)

LING539 – Statistical Natural Language Processing

Febin Jacob Varghese | febinv@email.arizona.edu

THE UNIVERSITY OF ARIZONA

Introduction

Twitter, a micro-blogging site has become a platform where millions of people express their opinions and react to events in real-time. Thus being able to process and mine this data can lead to interesting information about people, products, events etc.

Sentiment Analysis is a common technique used to analyze information from text and gather opinions. For e.g. the Obama administration had used it to gauge public opinion to policy announcements.

Task:

Given a tweet, classify it into one of seven ordinal classes, corresponding to various levels of positive and negative sentiment intensity, that best represents the mental state of the tweeter.

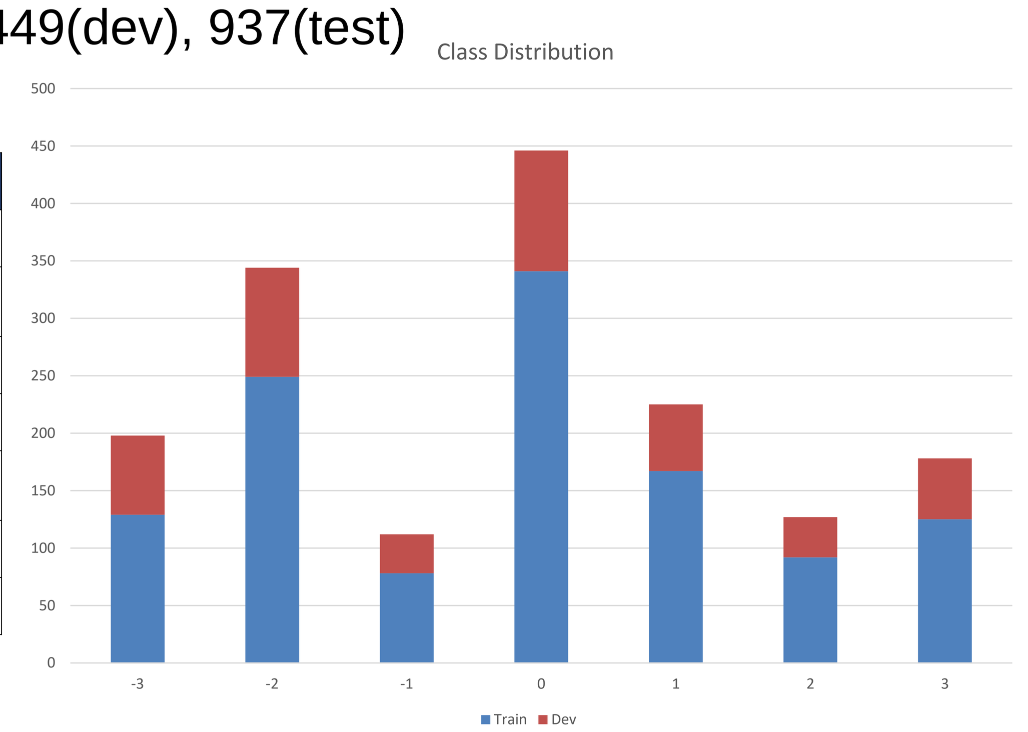
Challenges:

Unstructured text   Emojis   Sarcasm   Word Ambiguity   Negated Opinion

Twitter Evaluation Dataset

- Dataset: divided into Train, Dev and Test. Only considering English
- Labels: 7 label classes from -3 to +3
- #Records: 1181 (train), 449(dev), 937(test)
- Fields: ID , Tweet, Class

Intensity Class
-3: very negative emotional state can be inferred
-2: moderately negative emotional state can be inferred
3: very positive emotional state can be inferred
1: slightly positive emotional state can be inferred
2: moderately positive emotional state can be inferred
-1: slightly negative emotional state can be inferred
0: neutral or mixed emotional state can be inferred



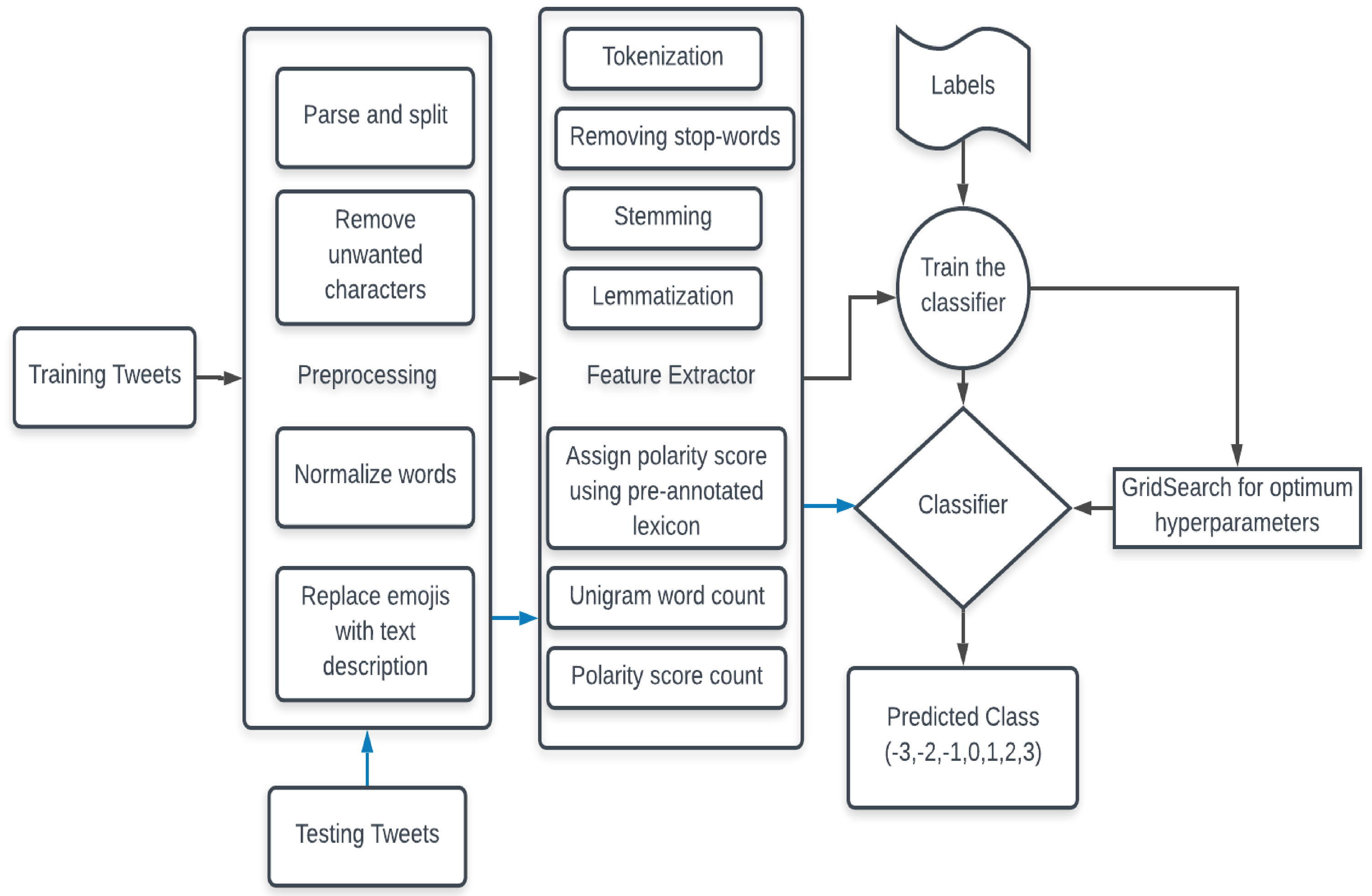
Pre-processing & Features

- Used Spacy package to perform lemmatization and tokenization
- Used nltk.VADER word lexicon to calculate polarity scores
- Used nltk.VADER emoji lexicon to replace emojis with text description of the emoji
- Removed unwanted characters and normalized tokens with urls, numbers, emails, username to generic keywords

Features	Source
Unigram Words	Extracted from tweet
Bigram Words	Extracted from tweet
Capitalized	Extracted from tweet
unigram lemmatized	Used Spacy's lemmatizer function on tweet
unigram stemmed	Used Spacy's lemmatizer function on tweet
positive polarity total score	Lookedup score using nltk.VADER package lexicon
negative polarity total score	Lookedup score using nltk.VADER package lexicon
compound polarity total score	summed positive and negative score calculated above for a tweet
count of positive words	count of positive words lookedup using vader lexicon
count of negative words	count of negative words lookedup using vader lexicon

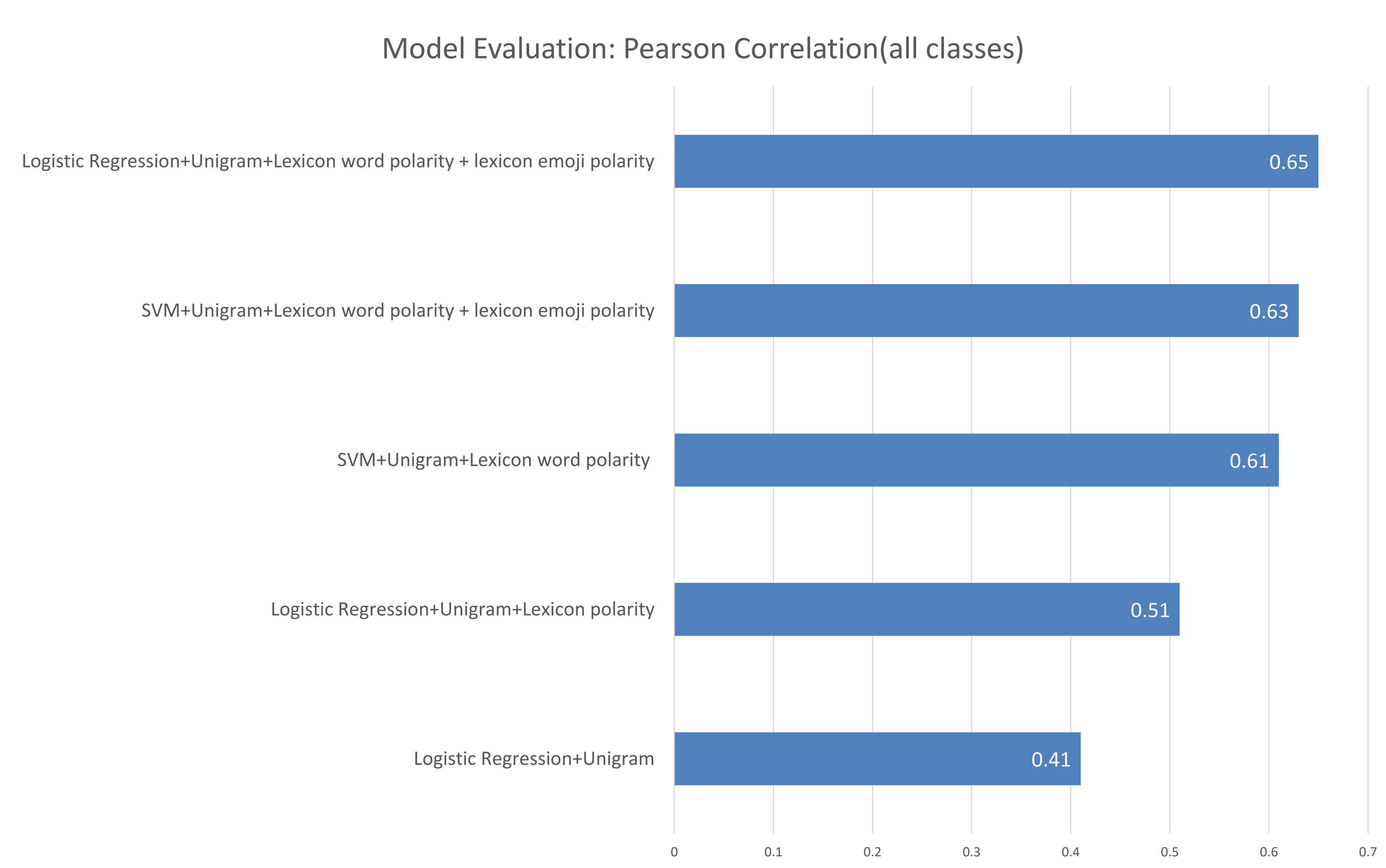
Approach

- Use a lexicon based approach to compute polarity scores along with other word based features to train a machine learning algorithm and perform prediction
- GridSearch was performed to infer the optimum hyperparameters for the model



Models Evaluated

- Two machine learning algorithms were evaluated: Logistic Regression and Support Vector Machines
- Best performing model was the logistic regression model



Results

SemEval 2018- Leaderboard

- Achieved 10th place on the leaderboard with the final model
- Total of 40 participants

4a. V-oc (valence ord.class.) English

#	User	Entries	Date of Last Entry	Team Name	Pearson (all classes)	Pearson (some-emotion)	Kappa (all classes)	Kappa (some-emotion)
					valence ▲	valence ▲	valence ▲	valence ▲
1	venkatesh-1729	1	02/06/18		0.836 (1)	0.884 (1)	0.831 (1)	0.873 (1)
2	danielfl	1	02/05/18	Amobee	0.813 (2)	0.865 (3)	0.812 (2)	0.863 (2)
3	psyml	1	02/01/18	psyML	0.802 (3)	0.869 (2)	0.796 (3)	0.859 (3)
4	anasteam	2	02/09/18		0.776 (4)	0.854 (4)	0.763 (5)	0.849 (4)
5	howard15	6	04/11/19		0.774 (5)	0.825 (5)	0.774 (4)	0.821 (5)
6	jogonba2	1	02/05/18	ELiRF-UPV	0.759 (6)	0.821 (6)	0.757 (6)	0.812 (6)
7	yoyo	1	02/02/18	YNU-HPCC	0.733 (7)	0.787 (7)	0.712 (7)	0.756 (7)
8	ISCLAB	1	01/29/18	ISCLAB	0.708 (8)	0.770 (8)	0.686 (8)	0.740 (8)
9	lanman	1	01/29/18	ECNU	0.686 (9)	0.747 (9)	0.684 (9)	0.740 (9)
10	febinjv	4	04/29/19		0.653 (10)	0.718 (10)	0.644 (10)	0.708 (10)
11	dragoni	5	02/12/18	NEUROSENT-PDI	0.331 (11)	0.370 (11)	0.301 (11)	0.323 (11)
12	utpalsikdar	1	02/05/18		0.164 (12)	0.178 (12)	0.098 (13)	0.099 (13)
13	NAVEEN.J.R	1	02/05/18	CENNLP	0.155 (13)	0.174 (13)	0.072 (14)	0.073 (14)
14	Amrita_student	1	02/06/18		0.125 (14)	0.140 (14)	0.117 (12)	0.127 (12)
15	thelonewolf190694	5	02/06/18		-0.004 (15)	-0.005 (15)	-0.001 (15)	-0.001 (15)
16	bicici	13	03/26/19	RTM	-999.999 (16)	-999.999 (16)	-999.999 (16)	-999.999 (16)

Findings, Limitations & Future Scope

Findings:

- Lexicon based polarity features performed better than unigram bag of words features. Removing stop-words, stemming and lemmatizing improved the performance
- Tri-gram features, Parts-of-speech, word shape, word/character length brought down the performance
- Best performing features were combination of lexicon based polarity features along with unigram and bigram bag of words

System Limitations:

- Error analysis of the model shows it has issues predicting sentiment where context, sarcasm, negation, slang words are involved
- Since lexicon based, has trouble capturing overall sentiment scores

Future Scope:

- Implement deep learning architectures such as LSTM's and CNN
- Implement BERT using pre-trained word and sentence embeddings such as GloVe

References

- *Venkatesh Duppada, Sushant Hiray "Seernet at EmoInt-2017: Tweet Emotion Intensity Estimator" arXiv:1708.06185 https://arxiv.org/abs/1708.06185*
- *Agarwal, Apoorv and Xie, Boyi and Vovsha, Ilia and Rambow, Owen and Passonneau, Rebecca "Sentiment Analysis of Twitter Data" http://dl.acm.org/citation.cfm?id=2021109.2021114*
- *Mansour R., Hady M.F.A., Hosam E., Amr H., Ashour A. (2015) Feature Selection for Twitter Sentiment Analysis: An Experimental Study. In: Gelbukh A. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2015. Lecture Notes in Computer Science, vol 9042. Springer, Cham*
- *https://spacy.io/*
- *https://www.nltk.org/_modules/nltk/sentiment/vader.html*