Sentiment Classification of Political Tweets

George Dill

Abstract

Sentiment Classification of sentences remains an unsolved problem in computer science. A number of supervised machine learning and preprocessing methods are tried on a labeled data set consisting of tweets about the 2012 presidential debates between Barack Obama and Mitt Romney. The data set is partitioned into tweets where the primary subject is Obama or Romney. Of the combinations of pre-processing and classification algorithms a Linear Support Vector Machine with an input vector of preprocessed text unigrams and bigrams weighted with TF-IDF produced the best results. The accuracy and F-score, in the low 60% for both the Romney and Obama data sets is unremarkable.

Introduction

In the era of social networking and crowd-sourced content online reviews can contain valuable insight for political analysts, product designers, retailers, and traders. However, systematically extracting aspects of an opinion document and the author's corresponding sentiment remains an unsolved problem.[1]

Sentiment classification is the determination of an author's "attitudes, thoughts, or judgement prompted by a feeling," as positive, negative, or neutral based upon the contents of the document.[2] Further complicating the problem is that an author may present multiple opinions, or "positive or negative view, attitude, emotion, or appraisal about an … aspect," with differing sentiments within the same document.[3]

This paper explores sentiment analysis of tweets, which are user generated posts to Twitter, an online forum where users, in 2012, could post messages no longer than 140 characters. Twitter messages can contain handles, represented by a leading '@' which alert the user following the '@' symbol of the tweet, hashtags, represented by a leading '#' which are user provided labels for the tweet, and also retweets, represented by a leading 'RT' symbolizing a source attribution to another user.

Tweets reduce the sentiment analysis from document level to sentence level because the 140-character limit generally restricts users to only one or two sentences.

The tweets analyzed in this paper were collected during the 2012 presidential debate between Barack Obama and Mitt Romney. The subject, or aspect, of the tweets has been pre-determined to be Barack Obama or Mitt Romney and the sentiment has been hand-labeled positive, negative, or neutral. This data set will represent the ground truth for our sentiment classifier.

We attempt to build a sentiment classifier which produces the best F-Score with the dataset given without overfitting the data.

---

[1] (Liu, CS583-Opinion Mining n.d.)
[2] (Liu 2011)
[3] (Liu, Web Data Mining 2011)

Techniques

Evaluation

All experiments in this paper were evaluated using 10-fold cross validation. 10-fold cross validation is an evaluation method where a data set is randomly partitioned into 10 subsets. The classifier is trained using 9 of the subsets and tested using the remaining test set.[4] A confusion matrix is generated for each combination and the accuracy, precision, recall, and F-score are calculated for each training set – test set combination. We report the average accuracy, precision, recall, and F-score across all trials scoring better than a random guess baseline classifier.

Classification Algorithms

We use off-the-shelf classification algorithms from SciKit-Learn, a machine learning library written in Python.[5] The classification algorithms used include:

Support Vector Machine – LinearSVC() in SciKit-Learn is finds a hyperplane between classes called a decision boundary that maximizes the margin, or distance from the hyperplane to the nearest point in space.[6] The SciKit-Learn implementation takes a strategy of one vs rest classification.[7] For our 3 classifications, 3 classifiers are trained Positive vs Other, Neutral vs Other, and Negative vs Other. We only use a linear kernel in this paper.

Multinomial Naïve Bayes – MultinomialNB() in SciKit-Learn considers each document a "bag of words" where word position is irrelevant. A document of length n is assumed to be generated by pulling words at random from a multinomial distribution. This assumption along with the conditional independence assumption allows a classifier to pick the max probability by the equation:

$$argmax_{c_j \in C} \Pr\left(c_j \mid d_i;\ \widehat{\Theta}\right)$$[8]

Laplace smoothing is implemented in this paper.

Decision Tree – A classification algorithm where the data set is greedily split into subsets using max decrease in entropy.

Random Forest – Many decision trees are generated at random, and with replacement, from the training data set. Each tree classifies the test value. The average classification is reported.[9]

Two Step Subjective and Polarity – We train two linear support vector machines, one to determine if a sentence is subjective or objective and another to determine its polarity. The subjective support vector machine was tested with two label sets: opinion word count greater than 1 and hand label of zero as the positive class. This method was inspired by Riloff and Wiebe.[10]

---

[4] (Liu, Web Data Mining 2011)
[5] (Pedregrosa 2011)
[6] (Liu, Web Data Mining 2011)
[7] (Pedregrosa 2011)
[8] (Liu, Web Data Mining 2011)
[9] (Pedregrosa 2011)
[10] (Riloff and Wiebe n.d.)

Counting Opinion Words – Hu and Liu's opinion lexicon is used to tally an opinion score based on the appearance of an opinion word from the tweet in the positive or negative lexicon.[11] Positive words are given a 1, negative words are given a -1. A sentence is classified positive if the sum is greater than 0, neutral if zero, or negative if less than 0.

Preprocessing

We examine a variety of preprocessing approaches and combinations thereof with classification algorithms.

Plain Text: Tweets are stripped of entity and attribute tags, leading handles, and hyperlinks. The '#' and '@' are removed from remaining hashtags and handles. All text is converted to lower case. The feature vector is built of both unigrams and bigrams. Both the TF-IDF and word count vectorization scheme are attempted.

Hashtags Only: A feature vector is built using only hashtags, weighted by TF-IDF and classified.

Handles Only: A feature vector is built using only handles, weighted by TF-IDF and classified.

Links Only: A feature vector is built using only hyperlinks.

Opinion Words: A tweet is scanned to see if it contains opinion words from Hu and Liu's lexicon of positive and negative words.[12] If the opinion score is non-zero a unique word is added to the end of the tweet to signify the positive or negative count.

Evaluation

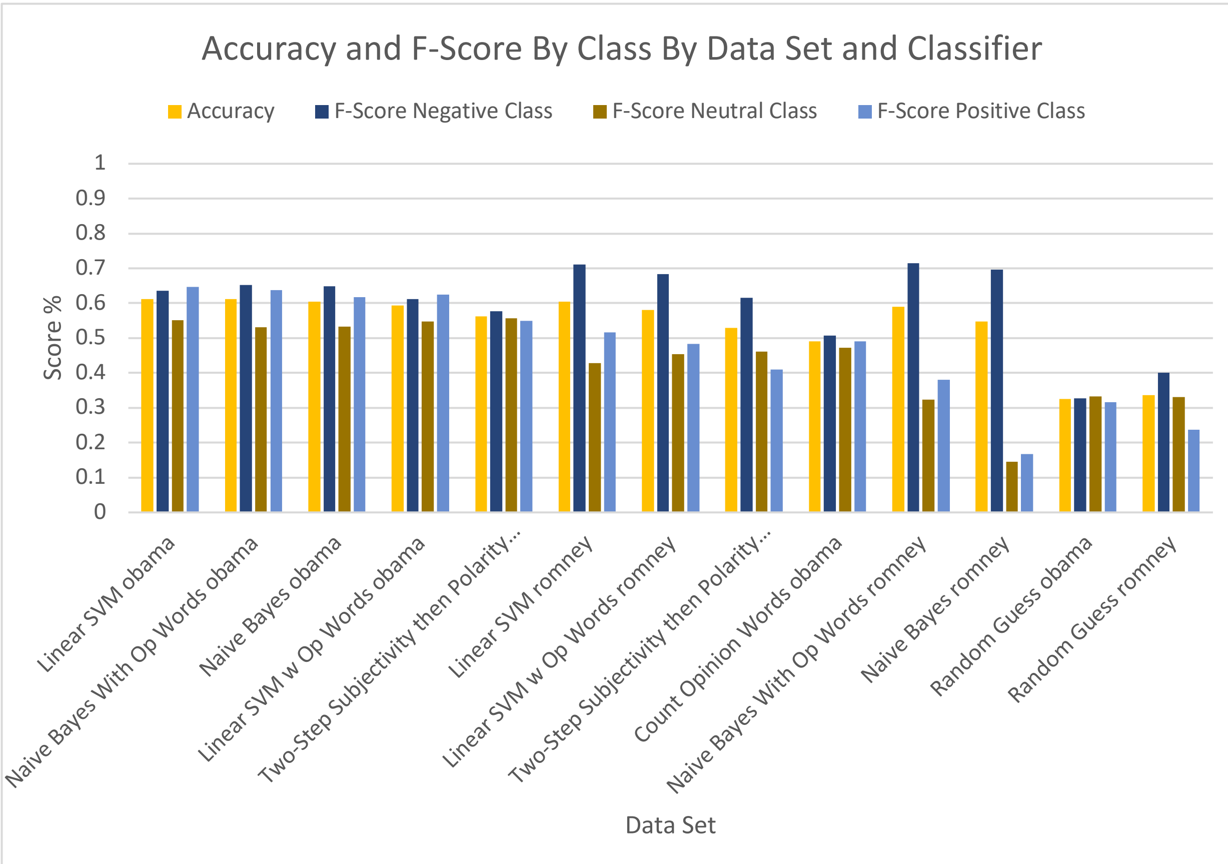
Appendix 1 reports the accuracy and F-score results for each classifier and feature set sorted by highest average F-score. Of the top 10 scoring classifiers we observe 8 take preprocessed text as the feature for the classifier. Table 1 summarizes the performance of the sentiment classifiers processing text in descending order of F-score.

We observe that the top-scoring classifier for both data sets is the Linear Support Vector Machine with the Obama set achieving accuracy of 61.1% and an average F-Score of 61.1% and the Romney set achieving accuracy of 60.4% and an average F-Score of 59.4%.

Surprisingly the 2-step classifier, which attempts to determine if a tweet is subjective or objective and then proceeds to classify sentiment on only subjective tweets was ranked 4th for the Obama dataset and 3rd for the Romney dataset with (accuracy, average F-score) of (56.1%, 56.1%) and (52.9%, 54.0%) respectively. Future work would include attempting to increase the accuracy of the Subjective/Objective classifier with a closer reading of (Hu and Liu 2004).

[11] (Hu and Liu 2004)
[12] (Hu and Liu 2004)

Table 1: Accuracy and F-Score By Data Set and Classifier



Conclusions

The classifier submitted for the class project will be the Linear SVM with text preprocessing as described above using unigrams and bigrams as the feature vector.

With accuracy in the low 60% sentiment analysis of political tweets is far from a solved problem. No analyst would care to use the results of this classifier to craft talking points of a candidate.

Although the 2-step classifier performed weaker than the SVM, it still provides opportunity to add new features to the classification and is worth exploring more in depth.

Works Cited

Hu, Minqing, and Bing Liu. 2004. "Mining and Summarizing Customer Reviews." *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Seattle, WA: ACM.

Liu, Bing. n.d. *CS583-Opinion Mining.* Accessed April 24, 2019. http://www.cs.uic.edu/~liub/teach/cs583-spring-19/CS583-link-analysis.ppt.

—. 2011. *Web Data Mining.* New York: Springer Heidelberg Dordecht.

Pedregrosa, et al. 2011. "Scikit-Learn: Machine Learning in Python." *JMLR 12* 2825-2830.

Riloff, Ellen, and Janyce Wiebe. n.d. "Learning Extraction Patterns for Subjective Expressions." *aclweb.* Accessed April 26, 2019. https://www.aclweb.org/anthology/W03-1014.

Classifier and Data Set	Feature	Accuracy	p neg	r neg	F-Score Negat	p neu	r neu	F-Score Neutr	p pos	r pos	F-Score Positiv
Linear SVM obama	text	0.611556308	0.60288198	0.673277711	0.635801612	0.587351027	0.519726004	0.550768888	0.648052038	0.647427619	0.647005291
Naive Bayes With Op Words obama	text	0.610992061	0.571563264	0.762991422	0.653053233	0.608990966	0.472845809	0.531369771	0.687199532	0.597620174	0.6376801
Naive Bayes obama	text	0.605160845	0.548614569	0.801194794	0.649226419	0.615532691	0.473307504	0.533494736	0.732287982	0.535814939	0.61723231
Linear SVM w Op Words obama	text	0.593958059	0.623098773	0.60111314	0.611292438	0.555657647	0.541065358	0.547790461	0.604082532	0.648569563	0.624941137
Two-Step Subjectivity then Polarity Classifier obama	text	0.561932364	0.614842677	0.545237365	0.57749494	0.480672716	0.660664408	0.556062303	0.673008673	0.465466526	0.549776132
Linear SVM romney	text	0.604110722	0.645143581	0.790278163	0.710169571	0.499745661	0.37670669	0.428585334	0.591828948	0.458933005	0.515841939
Linear SVM w Op Words romney	text	0.581270061	0.659613696	0.709050867	0.683103176	0.464582058	0.445063572	0.453639358	0.524210928	0.450928532	0.483323443
Linear SVM w Op Words obama	hashtags	0.502735601	0.577115314	0.498660304	0.53436556	0.448068285	0.495431976	0.470261474	0.498484889	0.516384344	0.506468628
Naive Bayes With Op Words obama	hashtags	0.501870398	0.571543656	0.498637493	0.531809279	0.446783744	0.493943817	0.468529204	0.50184358	0.513459083	0.506977178
Two-Step Subjectivity then Polarity Classifier romney	text	0.529027492	0.653544695	0.583419484	0.616130026	0.386682978	0.571596595	0.46047775	0.586481379	0.316657525	0.409435403
Naive Bayes With Op Words obama	callout	0.493147996	0.56023623	0.471728316	0.511671339	0.438894319	0.499961768	0.467212221	0.499735704	0.509658947	0.504167253
Linear SVM w Op Words obama	callout	0.492812761	0.5580134	0.465552419	0.507469326	0.438953213	0.50273709	0.468098407	0.500701704	0.511421776	0.505216601
Linear SVM w Op Words obama	links	0.493697677	0.571109305	0.456269795	0.507045772	0.432466521	0.605092834	0.503903556	0.532019856	0.406652107	0.459976971
Naive Bayes With Op Words obama	links	0.493681761	0.571554033	0.456373905	0.506654628	0.432413307	0.605581095	0.504273615	0.530259733	0.406289774	0.459808008
Count Opinion Words obama	text	0.489755025	0.570495022	0.455925997	0.506468595	0.435350589	0.518254771	0.472852312	0.489485492	0.494388329	0.491372724
Naive Bayes With Op Words romney	text	0.589776325	0.583826094	0.921949826	0.714333238	0.565273004	0.227898588	0.323750678	0.686047667	0.264507399	0.380669759
Linear SVM w Op Words romney	links	0.53523693	0.562022303	0.866940084	0.681774702	0.461336685	0.209384503	0.286875087	0.386086198	0.153954276	0.219012162
Naive Bayes With Op Words romney	links	0.532758767	0.562139371	0.866944452	0.681689246	0.454285259	0.215401328	0.290118349	0.33266589	0.128428855	0.18480199
Random Guess obama	hashtags	0.347197234	0.366732443	0.359512819	0.362925833	0.36827708	0.331766714	0.348914466	0.308232874	0.351210635	0.327878433
Naive Bayes romney	text	0.54833336	0.536902143	0.988161669	0.695488938	0.707332586	0.081560762	0.146040238	0.786555161	0.094560458	0.167189125
Random Guess obama	callout	0.333886884	0.35976382	0.342630826	0.350799	0.338075959	0.315925365	0.326145597	0.303837771	0.343823627	0.322149618
Random Guess obama	links	0.334230749	0.350509146	0.326085727	0.337364915	0.363770637	0.354296365	0.358798564	0.286591049	0.318553436	0.301104942
Random Guess obama	text	0.325172368	0.341122941	0.315288477	0.327012748	0.341180857	0.324939215	0.332130876	0.295737915	0.339137028	0.315547486
Random Guess romney	text	0.335684623	0.513983514	0.327800747	0.399991469	0.309326288	0.358519813	0.331728087	0.186034236	0.325100285	0.236282444
Random Guess romney	callout	0.331801506	0.515429317	0.328616658	0.400537566	0.295480167	0.339058217	0.315030304	0.18845591	0.331333251	0.239720291
Random Guess romney	hashtags	0.332677659	0.511833019	0.34889364	0.414399662	0.285482103	0.309031716	0.295767083	0.189212752	0.330023466	0.239609413
Random Guess romney	links	0.329857747	0.505841492	0.330662094	0.399561538	0.29994195	0.332600267	0.315272804	0.183846947	0.322926388	0.233602271
Naive Bayes obama	links	0.398045863	0.359182625	0.73162795	0.481573314	0.493776119	0.40352748	0.443824249	0	0	0
Linear SVM obama	links	0.398055412	0.358801698	0.730114813	0.480848745	0.493388849	0.404328879	0.444063378	0	0	0
Two-Step Subjectivity then Polarity Classifier obama	links	0.398025172	0.358726339	0.730396829	0.4807948	0.492877099	0.403624223	0.443392331	0	0	0
Linear SVM obama	hashtags	0.38754719	0.581475931	0.124876438	0.204907155	0.364496833	0.86849535	0.513292597	0.43576388	0.12969444	0.19841574
SGD romney	links	0.501782172	0.538859196	0.841223441	0.653099399	0.233870054	0.218725944	0.225679238	0.019135802	0.02605042	0.022064057
Linear SVM romney	callout	0.516294818	0.522952113	0.956195624	0.675893404	0.402714362	0.044526607	0.079919634	0.426788534	0.071246501	0.121696991
Two-Step Subjectivity then Polarity Classifier obama	hashtags	0.381715974	0.571902939	0.111878386	0.186092879	0.36169878	0.878937834	0.512236754	0.427020079	0.112356176	0.1765841
Two-Step Subjectivity then Polarity Classifier romney	hashtags	0.509915591	0.520633794	0.939652065	0.669529396	0.316079533	0.062256476	0.103620737	0.585681818	0.053039377	0.096586352
Linear SVM w Op Words romney	hashtags	0.516476839	0.523274514	0.958337055	0.676615776	0.366978477	0.044384493	0.078326818	0.485135402	0.065633772	0.114096599
Two-Step Subjectivity then Polarity Classifier romney	callout	0.512388437	0.523708106	0.948248831	0.674610833	0.348339546	0.053898395	0.093132115	0.403717627	0.055428525	0.097212821
Linear SVM w Op Words romney	callout	0.517164055	0.523740916	0.96251133	0.678048059	0.35437138	0.040140734	0.071956509	0.472160173	0.064683554	0.112218078
Naive Bayes With Op Words romney	callout	0.520533174	0.52437894	0.972541145	0.680955707	0.408339169	0.039126813	0.071196039	0.521868687	0.057738198	0.103120297
Naive Bayes With Op Words romney	hashtags	0.517527468	0.521786705	0.968746092	0.678038319	0.39517504	0.035903124	0.06521703	0.552878788	0.057310419	0.103461921
Linear SVM romney	hashtags	0.513636178	0.520253793	0.958623289	0.674260986	0.323030923	0.036910372	0.065586067	0.491149227	0.060880322	0.106819375
Naive Bayes obama	hashtags	0.372635077	0.507096623	0.365499305	0.298640009	0.368731876	0.652934217	0.395270314	0.560284784	0.085802319	0.14787648
Naive Bayes romney	hashtags	0.516817617	0.51774103	0.984433594	0.678228209	0.289996115	0.014289967	0.027101988	0.728080808	0.044367867	0.082764226
SGD obama	links	0.371206618	0.101142198	0.174581333	0.123534687	0.398890606	0.541908113	0.420331895	0.17140849	0.407543998	0.24125766
Linear SVM obama	callout	0.376168376	0.518590465	0.083315631	0.142945284	0.361853683	0.931635043	0.520731977	0.504517603	0.065985781	0.115566843
Naive Bayes romney	callout	0.515418035	0.51815049	0.984074517	0.678473816	0.387746698	0.028400698	0.052612192	0.591666667	0.015829136	0.030669063
SGD romney	hashtags	0.518603247	0.517897813	0.996852727	0.681461045	0	0	0	0.645357143	0.043116307	0.079946494
Naive Bayes obama	callout	0.361768214	0.489030167	0.251406977	0.210754454	0.373055963	0.752906444	0.425222891	0.529774436	0.0672241	0.119096697
Two-Step Subjectivity then Polarity Classifier obama	callout	0.372979176	0.523317156	0.07444612	0.129798717	0.36045555	0.936767117	0.520245693	0.479723942	0.059144064	0.104936515
SGD obama	hashtags	0.351824901	0.404246419	0.616779293	0.402282127	0.420055594	0.11955402	0.11064256	0.421565757	0.332919354	0.239515084
SGD obama	callout	0.354522082	0.508360421	0.256834144	0.221454827	0.426468142	0.400453651	0.264959734	0.417675659	0.417927258	0.259341247
SGD romney	callout	0.512050174	0.513841051	0.995199272	0.677614106	0.364285714	0.002955919	0.005836829	0.144444444	0.00740985	0.014022755
Linear SVM romney	links	0.512222135	0.512222135	1	0.677312446	0	0	0	0	0	0
Naive Bayes romney	links	0.512216476	0.512216476	1	0.677182135	0	0	0	0	0	0
Two-Step Subjectivity then Polarity Classifier romney	links	0.512217105	0.512217105	1	0.677144963	0	0	0	0	0	0