

Lexical and Machine Learning approaches toward Online Reputation Management

Chao Yang, Sanmitra Bhattacharya, and Padmini Srinivasan

Department of Computer Science,
University of Iowa, Iowa City, IA, USA
{chao-yang, sanmitra-bhattacharya, padmini-srinivasan}@uiowa.edu

Abstract. With the popularity of social media, people are more and more interested in mining opinions from it. Learning from social media not only has value for research, but also good for business use. RepLab 2012 had Profiling task and Monitoring task to understand the company related tweets. Profiling task aims to determine the Ambiguity and Polarity for tweets. In order to determine this Ambiguity and Polarity for the tweets in RepLab 2012 Profiling task, we built Google Adwords Filter for Ambiguity and several approaches like SentiWordNet, Happiness Score and Machine Learning for Polarity. We achieved good performance in the training set, and the performance in test set is also acceptable.

Keywords: Polarity, Ambiguity, Company, Twitter, SentiWordNet, Happiness Score, Google Adwords

1 Introduction

Social media has become an integral part of our everyday life. The increasing influence of social media on our daily life can be observed in various scenarios, ranging from gathering movie reviews [1] to understanding health beliefs[2] Given the number of online users voicing their personal opinion on various topics on social media streams such as Twitter, it's now feasible to aggregate opinions of the public to create meaningful inferences. Reputation management is one such area where public opinion towards a topic (such as a company or product) is aggregated. Traditional methods of reputation management are essentially based on word-of-mouth or surveys which are not only expensive but also time consuming. With the advent social media, reputation management can be done rapidly, more extensively and at a cheaper cost. The "evaluation campaign for Online Reputation Management Systems" or RepLab 2012, aimed towards this goal of aggregating public views on a company to see how a company (or it's products) are perceived among online users. The goal was also to gauge company strengths and weaknesses and most importantly, from the company's perspective, predict early threats to it's reputation and thereby neutralize them before they become widespread. Keeping this in mind, Replab 2012 had two tasks: Profiling task and Monitoring task using Twitter data.

Our group only participated in the Profiling task. Here systems were required to automatically work on two different aspects related to tweets: Ambiguity and Polarity. For Ambiguity one needs to judge if there is a relationship between the tweet and the company. For example, in the tweet “Apple May Legally Force Motorola To Destroy Their Phones <http://nblo.gs/uFvFb>”, ‘apple’ refers to the company Apple, Inc. On the other hand, in the tweet “I need to get off the coffee and eat my apple and carrots.”, ‘apple’ is a fruit. In this task, a tweet needs to be judged as relevant or irrelevant w.r.t. a company name. Polarity of a tweet is defined as the polarity w.r.t. the reputation of a company. For instance, the tweet “Lufthansa announces major expansion in Berlin with opening of new Brandenburg Airport in June 2012” entails a positive view towards the company ‘Lufthansa’, and hence has a positive influence on the company’s reputation. On the contrary, the tweet “#Freedomwaves - latest report, Irish activists removed from a Lufthansa plane within the past hour.” entails a negative view towards ‘Lufthansa’ and hence it may have negative influence on same company’s reputation. As a third category, there can also be tweets that have neither positive nor negative influence on a company’s reputation (e.g. “I’m at Lufthansa Aviation Center (LAC) (Airportring 1, Frankfurt am Main) w/ 2 others <http://4sq.com/vTCDiA>”). Such cases are identified as neutral. Participating systems are required to declare each tweet as positive, negative or neutral w.r.t. a company’s reputation.

This paper describes our five run submissions. Run 1 and Run 2, we use the popular sentiment lexicon, SentiWordNet [3], to identify the polarity of tweets. To determine the ambiguity of tweets, we use a Google AdWords¹ filter in Run 1, while in Run 2 we treat all tweets as relevant to some companies. For Run 3 and Run 4, we used a *Happiness lexicon* (discussed later) to identify the polarity of tweets. Similar to Run 1, Run 3 also uses Google AdWords filter to judge the ambiguity tweets, while Run 4 treats all tweets as relevant to some companies. Finally, for Run 5, we again treat all the tweets as relevant to some companies, and use a machine learning approach to classify the polarity of test tweets. The classifier was built using all the tweets in training set.

Below we describe our Google AdWords filter for judging ‘Ambiguity’, SentiWordNet and ‘Happiness lexicon’-based approaches for polarity, and lastly our classifier-based approach for polarity. We conclude with a discussion of the performance of our submitted runs.

2 Dataset Acquisition

Similar to the TREC Microblog track datasets, tweets about the various companies were not distributed directly due to Twitter’s data sharing policies, but participating teams were given tweet IDs and associated information for accessing the tweets directly from Twitter. Tools for downloading the tweet contents from the Twitter servers were provided. However, this task proved to be challenging. Since Twitter data is dynamic (users may delete tweets or even delete

¹ <https://adwords.google.com/>

accounts), this resulted in differences in datasets collected by the different participating research groups at different times.

3 Dataset Statistics

The training set comprised of 300 tweets each for 6 companies. Table 1 shows the statistics of the training set.

Company	apple	lufthansa	alcatel	armani	barclays	marriott
Total Tweets	300	300	300	300	300	300
Relevant	281	299	289	179	298	294
Non-relevant	19	1	11	21	2	6
Null Tweets	33	37	32	92	16	46
English	74	228	236	270	292	285
Spanish	226	72	64	30	8	15
Positive	70	242	221	24	248	94
Neutral	195	35	64	155	24	192
Negative	18	20	4	5	27	11

Table 1. Statistics of Training Set

The null tweets are the ones which could not be obtained from Twitter because of the aforementioned problem (Section 2). In the training set, most of the tweets are relevant to the company. Armani has the most non-relevant tweets which are only 21. English tweets dominate the training set except the tweets for Apple. We also observed the there are not too many negative tweets for each company.

4 Methods

4.1 Google AdWords Filter For ‘Ambiguity’

Analysis of Tweets in Training Set Before developing methods to determine the ambiguity of tweets we manually look through the tweets in training set. Basically, we found the tweets for ‘Apple’ are judged as relevant or non-relevant mostly by the following factors shown in Table 2.

This table shows the various factors we found in training set that could differentiate the ‘Ambiguity’ for Apple Inc. So our hypothesis for determining ‘Ambiguity’ for a company is: if a tweets has one or more keywords related to company factors, it is labeled as relevant. Otherwise, it is labeled as non-relevant.

It is true that different companies have different factors. But generally, there are some common company factors: specific products, generic products, competitors, official tweet account name, company name hashtag, company leaders, and official website.

	Factors	Example
Apple as Company	Specific products	iTunes, Apple TV, iPad, etc.
	Generic products	phone, tablet, etc.
	Apps & Music	Angry Birds, etc.
	Competitors & their products	Samsung, HTC, etc
	Leader name	Steve Jobs
	Company related term	products, trademark dispute
Apple as Fruit	Verb.	eat
	Detail fruit related term	apple sauce, apple soup, pie, etc.
	Generic fruit related term	fruit, food, etc.
	Other fruits	banana

Table 2. Factors to Differentiate ‘Ambiguity’

Automatically acquiring company factors for each company is not a trivial task. In this paper we propose a new method for getting the company factors, using Google AdWords Keyword Tool[4]. Google AdWords Keyword Tool is a service from Google to help advertisers choose a few search terms related to their business. The keywords are the most frequent searched words by the internet users. Using the Keyword Tool, we can easily get the popular products for the company, generic products, and other company related terms.

Table 3 shows the top 20 out of 787 English Google AdWords for Apple, Inc. collected on Jun. 6th 2012 .

Rank	Keyword	Rank	Keyword
1	apple	11	apple iphone 8gb
2	apple store	12	apple iphone 5
3	apple iphone 4	13	apple iphone case
4	apple support	14	apple i4
5	apple iphone 4g	15	apple iphone covers
6	apple iphone	16	apple iphone 4gs
7	apple ipod touch	17	apple website
8	apple iphone support	18	apple 3g iphone
9	apple 3g	19	apple bumper case
10	apple i phones	20	apple 4g phone

Table 3. Top 20 Google AdWords for Apple Inc.

Although there are some keywords maybe mentioned the same product, for example ‘apple iphone’, ‘apple i phones’. But in total number of 787 keywords, they covered most of the popular products and Apple Inc. related keywords.

We developed two strategies to match the keywords. In the first strategy (Filter 1) we determine if a tweet has a whole keyword or not. In the second strategy (Filter 2) we determine if the tweet has one or more tokens of the

keywords. Both strategies we exclude company's name in the keywords. For example, 'iphone 4' is a keyword for Apple Inc. The tweet 'I love iphone', it doesn't contain the whole keyword. Hence by Filter 1 it is considered irrelevant while by Filter 2 it considered relevant.

Flow Chart of Google AdWords Filter Figure 1 shows the flowchart of our Google AdWords Filter. For one company, we use the company name or website as the search query to extract both English and Spanish keywords. So we have 4 lists of keywords for one company. Then we merge them into one large list, and also automatically add the Twitter account and hashtag for this company. For example, we can add '@apple' and '#apple' as the official account and hashtag for Apple Inc. Finally, if the tweet has one and more keywords, it is labeled as relevant.

4.2 SentiWordNet Approach For Polarity

Although polarity for reputation is substantially different from standard sentiment analysis, it does have some similarity. The polarity of a tweet is, in essence, expressed using certain sentiment-loaded keywords. For instance, in the tweet 'Lehmann Brothers goes bankrupt', the word 'bankrupt' has a negative polarity for reputation. Thus, if we can determine the polarity for each word in tweet, we may be able to determine the polarity of tweet.

SentiWordNet Since there is no specific polarity scores list for words to determine the polarity of reputation, so we decided to use SentiWordNet to get the polarity of individual words. SentiWordNet is a lexical resource for sentiment analysis and opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity. That is, SentiWordNet has a list of negative and positive scores for words. One word may have several negative and positive scores in different cases. Table 4 shows an example of 'bankrupt'

POS & ID	Word	PosScore	NegScore
n 09838370	bankrupt#1	0	-0.625
v 02318165	bankrupt#1	0	0

Table 4. Example of 'bankrupt' in SentiWordNet

The pair (POS & ID) uniquely identifies a WordNet (3.0) synset. Because the word 'bankrupt' has 2 cases in SentiWordNet, we calculate an average *PosScore* and *NegScore* for this word.

To determine the polarity for the tweet we use two strategies. Both strategies assign an *polarity score* for a tweet, then use two thresholds to determine the polarity of the tweet.

In the first strategy (MaxS) we use the maximum SentiWordNet scores among all the words to determine the polarity of tweets. Here we find the maximum *PosScore* and *NegScore* for each word of the tweet first. For example, in the tweet ‘Lehmann Brothers goes bankrupt’, after excluding the company name, ‘bankrupt’ has the maximum *NegScore* and ‘goes’ has the maximum *PosScore*. So the sum of these scores are the *polarity score* for the tweet.

In the second strategy (SumS) we use the sum of SentiWordNet scores of all the words. For example, again in ‘Lehmann Brothers goes bankrupt’, after excluding the company name, the sum of *PosScores* and *NegScores* of ‘goes’ and ‘bankrupt’, becomes the *polarity score* of the tweet.

Thus, both of the strategies give *polarity score* for each tweet. We then set two fixed thresholds: positive threshold and negative threshold. If the *polarity score* is larger than positive threshold, we claim the tweet is positive. If the *polarity score* is smaller than negative threshold, we claim the tweet is negative. Otherwise, the tweet is neutral. For Spanish tweets, we use Google Translate² to translate all the words in SentiWordNet to Spanish words. Then we use the translated Spanish SentiWordNet to determine the polarity of Spanish tweet.

4.3 Happiness Score Approach For Polarity

In addition to the SentiWordNet score, we also tried another approach to determine the polarity of words in a tweet. *Happiness score*[5] which developed by Dodds et al by crowdsourcing. It has a list of words. Each word is associated with a score to indicate the happiness of this word.

We use similar strategy with MaxS, using *Happiness scores* instead of using SentiWordNet scores. Firstly, we get the maximum *Happiness score* among all the tokens in tweet. Then we denote it as the *polarity score* for this tweet. Finally we set two threshold to determine the polarity of the tweet. We denote this approach as HappyS.

4.4 Machine Learning Approach For Polarity

Instead of SentiWordNet and *Happiness score* approaches for polarity, we also developed a machine learning approach. Figure 2 shows the flowchart of Classifier. We take all the labeled tweets from the 6 companies ($6 * 300 = 1800$ tweets) and search for the Google AdWords and replace them with ‘xxxx’, also replace the company names with ‘aaaa’. Then split this set into 2 parts by English and Spanish tweets. Build 2 separate classifiers one for English and other for Spanish. To evaluated the performance of the Classifier Approach, we train the classifiers using the training tweets from 5 companies and test on the remaining company tweets in the training set. Repeat this 6 times such that it covers all possibilities.

Especially, we use Weka[6] to test the performance of classifier. We use Bagging[7] classifier with SVM[8] kernel (SMO[9] in weka). We extract unigram, bigram and 3-gram features for the classifier.

² translate.google.com/

5 Results

5.1 Performance of Google AdWords Filter in Training Set

Tables 5 and 6 show the performance of Filters 1 and 2 on the training set.

	Apple	Lufthansa	Alcatel	Armani	Barclays	Marriott	Avg.
Precision	0.9785	0.9955	1	0.9572	1	1	
Recall	0.8114	0.7425	0.9204	0.9040	0.9161	0.6938	
F-score	0.8872	0.8506	0.9586	0.9299	0.9562	0.8193	0.9003

Table 5. Performance of Filter 1 on Training Set

	Apple	Lufthansa	Alcatel	Armani	Barclays	Marriott	Avg.
Precision	0.9710	0.9963	1	0.9171	1	1	
Recall	0.9537	0.9097	0.9516	0.9274	0.9664	0.7891	
F-score	0.9623	0.9510	0.9752	0.9222	0.9829	0.8821	0.9460

Table 6. Performance of Filter 2 on Training Set

We used precision, recall and F-score to evaluate the performance on the training set. Especially, we use average F-score (avgF) to evaluate the overall performance. We see that Filter 2 has better performance with not only higher avgF, but also all F-scores for every company are higher except for Armani. The reason is because, as expected, the Google Adwords keywords do not cover all the company factors. For example, ‘ios 5’ is a Adwords keyword for Apple Inc, but ‘ios’ is not. Filter 2, which favors more relaxed filtering covers more cases resulting in better performance. Therefore, we use Filter 2 for our submitted runs on the test set.

5.2 Performance of SentiWordNet and Happiness Score Approach in Training Set

Table 7 shows the performance of MaxS, SumS and HappyS.

avgF is the average F-score of Positive F-score, Neutral F-score, Negative F-score for the six companies in training set. The thresholds are automatically generated by script.

In the table, MaxS has better performance in English tweets. While HappyS has better performance in Spanish tweets. So we decided to use MaxS and HappyS in the submit runs. In addition, we can see that all the avgF is just around 0.3, which indicate the polarity task is difficult.

	MaxS		SumS		HappyS	
	En	Es	En	Es	En	Es
avgF	0.34	0.31	0.336	0.30	0.27	0.35
Positive threshold	0.62	1.26	0.34	4.08	23.4	27.3
Negative threshold	-0.3	-0.83	-2.59	-1.11	16.1	16.7

Table 7. Performance of MaxS and SumS in Training Set

5.3 Performance of Machine Learning Approach in Training Set

Table 8 shows the performance of the classifier.

	Classifier (En)		Classifier (Es)	
	unigram	3-gram	unigram	3-gram
avgF	0.264	0.269	0.259	0.273

Table 8. Performance of Classifier Approach in Training Set

Basically, Classifier Approach performance worse than SentiWordNet and Happiness score approach. None of the 4 avgF is above 0.3. We also tried using the webpage content linked in tweet, but the performance get worse. The reason is that the polarity of tweets can be expressed by so many ways, we need much more training tweets in order to get better performance.

Thus we denote classifier using 3-gram feature as Cla3. And use it in our submitted runs.

6 Submit Runs Description

We submitted five runs which has different combination of our ‘Ambiguity’ Approach and Polarity Approach. We denote ‘AllRel’ as treat all the tweets relevant for some companies. Table 9 shows the detail of the five different runs

Run ID	Description
Run1	Filter2 + MaxS
Run2	AllRel + MaxS
Run3	Filter2 + HappyS
Run4	AllRel + HappyS
Run5	AllRel + Cla3

Table 9. Description of Submitted Runs

7 Performance of Test Set

We describe the performance of test set separately by Filtering (‘Ambiguity’) and Polarity tasks. In Filtering task, our Run1 and Run 3 ranked in 7th and 8th place out of 33 runs, (5th of 9 teams) by F(R,S) score[10]. R and S refers Reliability and Sensitivity respectively. Because we treat all the tweets relevant in Run2, Run4, and Run5. They have the same results with the baseline (all relevant).

Table 10 shows the top 10 results of Filtering task.

Run ID	F(R,S)	R	S	Accuracy
replab2012_related_Daedalus_2	0.263922126	0.243482396	0.432991032	0.722763591
replab2012_related_Daedalus_3	0.253463929	0.235162463	0.422129397	0.702232013
replab2012_related_Daedalus_1	0.250619268	0.23968238	0.403657787	0.718006364
replab2012_related_CIRGDISCO_1	0.227595261	0.217923478	0.336440429	0.701870318
replab2012_profiling_kthgavagai_1	0.222829043	0.253419399	0.357636447	0.774061038
replab2012_profiling_OXY_2	0.196601614	0.234666227	0.272356458	0.809025193
replab2012_profiling_uiowa_1 (Run1)	0.177919294	0.181556704	0.292220139	0.679680848
replab2012_profiling_uiowa_3 (Run3)	0.177919294	0.181556704	0.292220139	0.679680848
replab2012_profiling_ilps_4	0.15730978	0.157010828	0.223508777	0.599100149
replab2012_profiling_ilps_3	0.155698416	0.155160491	0.25552382	0.657567983

Table 10. Top 10 Runs of Filtering Task by F(R,S) Score

One of the reason the test set results is not as good as training set is because some of the companies in test set do not have ambiguity. For example, if a tweet has Google or Microsoft in it. It definitely relevant with the company it mentioned. The right thing to do is to determine if the company has ambiguity. If not, we should treat all the tweets as relevant.

In Polarity task, our Run2 ranked in 4th place out of 31 runs, (4th of 9 teams) by F(R,S) score. The result shows using SentiWordNet approach is better than *Happniess score* and Classifier approach.

Table 11 shows the top 10 results of Polarity task and all of our other Runs.

8 Conclusion

In RepLab 2012, we explored using Google AdWords as a filter to determine the ambiguity of tweets. We also developed several approaches like SentiWordNet, Happiness Score, Classifier to determine the polarity of tweets. The results in test set shown our approaches performed well. However, our research still has some limitations. Google AdWord does provide great company related keywords, but it is not free service. We didn’t received approve from Google to use AdWord API before submitting the result. So we manually downloaded the English and Spanish keyword list searched by company name as query. The limitation of queries

Rank	Run ID	F(R,S)	R	S	Accuracy
1	replab2012_polarity_Daedalus_1	0.401818195	0.392370769	0.449091977	0.479550085
2	replab2012_profiling_uned_5	0.341946295	0.340229898	0.374731432	0.449501547
3	replab2012_profiling_BMedia_2	0.341946295	0.340229898	0.374731432	0.449501547
4	replab2012_profiling_uiowa_2 (Run2)	0.341946295	0.340229898	0.374731432	0.449501547
5	replab2012_profiling_uned_2	0.341946295	0.340229898	0.374731432	0.449501547
6	replab2012_profiling_uned_4	0.341946295	0.340229898	0.374731432	0.449501547
7	replab2012_profiling_BMedia_3	0.341946295	0.340229898	0.374731432	0.449501547
8	replab2012_profiling_OPTAH_1	0.341946295	0.340229898	0.374731432	0.449501547
9	replab2012_profiling_OPTAH_2	0.341946295	0.340229898	0.374731432	0.449501547
10	replab2012_profiling_BMedia_5	0.341946295	0.340229898	0.374731432	0.449501547
19	replab2012_profiling_uiowa_1 (Run1)	0.255176622	0.315109492	0.249941079	0.274533823
23	replab2012_profiling_uiowa_4 (Run4)	0.240957995	0.264677783	0.249820237	0.397726112
26	replab2012_profiling_uiowa_5 (Run5)	0.211165461	0.375737392	0.177001887	0.425064303
30	replab2012_profiling_uiowa_3 (Run3)	0.150727485	0.231986766	0.139879816	0.321687051

Table 11. Top 10 Runs of Polarity Task by F(R,S) Score

lead the limitation of the AdWord keywords. Then the performance of ‘Ambiguity’ Filter is limited. Another limitation is that SentiWordNet is a general list for the sentiment of words, it is not optimized for determining the polarity of companies. For example, ‘Expand’ is a positive word for judging polarity. But it is almost neutral in SentiWordNet. Thus exploring Google AdWords API query to get more company related keywords, and customizing a new polarity word list based on SentiWordNet could be the future work.

References

1. Asur Sitaram, Huberman Bernardo A. Predicting the Future with Social Media. March 2010.
2. Bhattacharya Sanmitra, Tran Hung, Srinivasan Padmini and Suls Jerry. Belief Surveillance with Twitter. In Proceedings of the Fourth ACM Web Science Conference (WebSci12), pages 5558, Evanston, IL, USA, 2012.
3. Esuli Andrea, Sebastiani Fabrizio. Sentiwordnet: A Publicly Available Lexical Resource For Opinion Mining. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06, pages 417422, 2006.)
4. Google Adwords Keyword Tool. URL <http://support.google.com/adwords/bin/answer.py?hl=en&answer=147602>.
5. Dodds P.S., Harris K.D., Kloumann I.M., Bliss C.A. and Danforth C.M. Temporal Patterns of Happiness and Information in A Global Social Network: Hedonometrics and Twitter. PLoS ONE, 6(12):e26752, 2011.
6. Hall Mark, Frank Eibe, Holmes Geoffrey, Pfahringer Bernhard, Reutemann Peter and Witten Ian H. The Weka Data Mining Software: An Update. SIGKDD Explorations, 11(1), 2009.
7. Breiman L. Bagging Predictors. Machine learning, 1996.
8. Vapnik V.N. The Nature of Statistical Learning Theory. HeidelbergA, (1995).

9. Platt John C. Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In B. Schoelkopf and C. Burges and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, 1998.
10. Amigo Enrique, Gonzalo Julio and Verdejo Felisa. Reliability and sensitivity: Generic Evaluation Measures For Document Organization Tasks. Technical report, Departamento de Lenguajes y Sistemas Informaticos, UNED, Madrid, Spain, 2012.

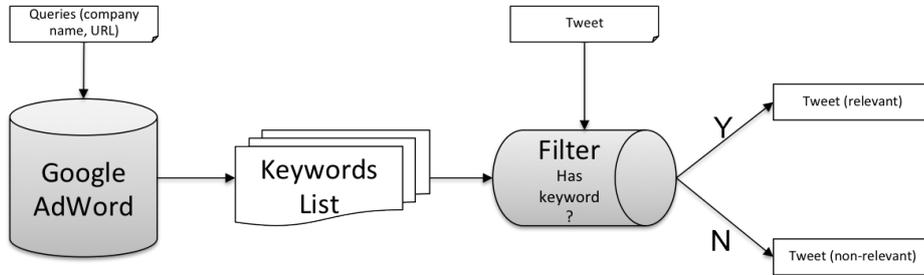


Fig. 1. The Flow Chart for Google AdWords Filter

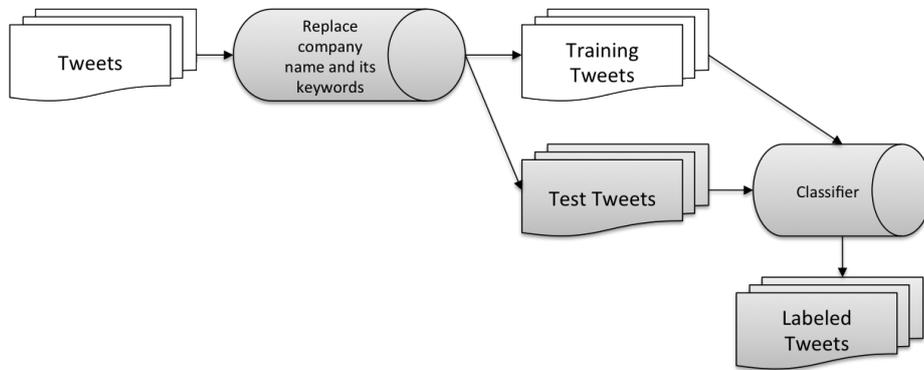


Fig. 2. The Flow Chart for Classifier For Polarity