

# Product Reputation Trend Extraction from Twitter

**Aizhan Bizhanova, Osamu Uchida**

Graduate School of Engineering, Tokai University, Hiratsuka, Japan  
Email: [aizhan.bizhanova@gmail.com](mailto:aizhan.bizhanova@gmail.com), [o-uchida@tokai.ac.jp](mailto:o-uchida@tokai.ac.jp)

Received 10 May 2014; revised 16 June 2014; accepted 7 July 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Micro-blogging today has become a very popular communication tool among the Internet users. Real-time web services such as Twitter allow users to express their opinions and interests, often expressed in the form of short text messages. Many business companies are looking into utilizing these data streams in order to improve their marketing campaigns, refine advertising and better meet their customer needs. In this study, we focus on using Twitter, for the task of extraction product reputation trend. Thus, business could gauge the effectiveness of a recent marketing campaign by aggregating user opinions on Twitter regarding their product. In this paper, we introduce an approach for automatically classifying the sentiment of Twitter messages toward product/brand, using emoticons and by improving pre-processing steps in order to achieve high accuracy.

## Keywords

Twitter, Reputation Analysis, Sentiment Analysis, Natural Language Processing

---

## 1. Introduction

Micro-blogging today has become a very popular communication tool among the Internet users. Millions of messages are appearing daily in popular websites that provide services for micro-blogging such as Twitter, Tumblr, Facebook. Authors of those messages write about their life, share opinions on variety of topics and discuss current issues. Such data can be efficiently used for marketing or social studies [1].

Through these opinions, we can extract information about the product, that we are interested in and numerate reputation of product. Knowing the reputation is very important for marketing analyzer because they enhance the public's view of product by analyzing extracted reputation. In the past, market analyzer conducted manual survey to find reputation of product. However, manual survey not only costs high but also requires lots of labor.

The purpose of our study is to extract opinion from micro-blog automatically and to summarize extracted opinions to provide reputation of product in which we are interested. In most of the previous researches, text polarity were extracted based on assumption that most of sentiment messages consists of positive or negative words as “good”, “bad” and etc. However, Twitter message’s structure is unique and it allows you to write messages no longer than 140 characters, which constrain users not to use very long sentences but to use emoticons, abbreviations, acronyms and other forms of informal language. Most of social networks users use informal language to shorten their messages, as it takes less time to type. So considering that, in presented research, we make an assumption that using emoticons, emotion identifiers, acronyms and etc. as sentiment classification feature that would help us to get high accuracy in sentiment extraction task.

In this paper, we propose a method to extract sentiment automatically from tweets, which are the Twitter user’s status messages. Many companies want to analyze their customer satisfaction, thus we apply our method to a “negative” and “others” (positive and objective tweets) classification task of tweets. We assume that “negative” tweets can be more informative, so merchandise department can use it to gather critical feedback about problems in newly released products.

## 2. Existing Study

Multiple papers have been published on sentiment analysis. Many of them have also explored using Twitter as their primary source of data.

Earlier works on sentiment analysis uses the traditional text classification methods on normal text forms like movie reviews. In [2], authors present a comprehensive comparison of machine learning algorithms in a fairly narrow domain of film reviews. Starting from being a document level classification task it has been handled at the sentence level [3] and more recently at the phrase level [4]. These methods are mainly fully supervised [5] which uses manually labeled data to train the classifier. Recently, distantly supervised methods [6], as using emoticons as noisy labels, and integration of these two methods [7] into the same learning framework were proposed. In the work of [8], Agarwal *et al.* are examining sentiment analysis on Twitter by conducting experiment with unigram model, a feature based model and a tree kernel based model.

In our work, we will pay attention to the most important pre-processing step before training the classifier. Emoticons, which can give us a lot of information about text sentiment are usually ignored or stripped as noisy labels. Thus, we believe that, by using emoticons in text sentiment classification we can get high accuracy in performance of our classifier.

## 3. Approach

Our approach is to use Naïve Bayes machine learning classifier for sentiment classification. First, we present how we collect data for training and test set. Then, we propose a very effective and efficient way of tweets pre-processing. Finally, we will present the results of experiment.

### 3.1. Data Gathering

In this work for tweets collection, Twitter API [9] was used. Twitter is an information network and communication mechanism that produces more than 300 million tweets per day [7]. The Twitter platform offers access to that corpus of data, via APIs. The Twitter API supports searching tweets pertaining to a query, thus we can obtain a large training set.

In this study, to collect data for each class (“negative” and “others”, as for “others” class we use “positive” + “neutral” tweets), positive (“☺”) and negative (“☹”) emoticons were used. As for neutral/objective tweets, spam or commercial tweets about product or service were considered as objective. We also make an assumption, that most of positive tweets toward product or service must contain positive expression words, like “good”, “great”, “amazing”, when words like “bad”, “awful” describes negative feelings. Thus, we increased our training set with tweets, which contains feeling descriptive words [10].

### 3.2. Data Pre-Processing

Twitter users are much more likely to have grammatical/spelling errors, colloquialisms, and slang incorporated into their output, due to the 140 character limit that is imposed on users. As a result, regular expression matching

of common errors and substituting with standard language is necessary.

In this study we introduce new resources for pre-processing Twitter data:

- 1) We replaced all emoticons with their sentiment polarity by looking up to the emoticon dictionary [11]. In **Table 1**, we show part of emoticons, from our emoticons dictionary, with its replacing pattern.
- 2) Non-informative Twitter usernames, URL links and hash tags were stripped from the tweets.
- 3) We build an acronym dictionary, to replace acronyms as OMG (“Oh My God”), LOL (“Laughing Out Loud”), ILU (“I Love You”) and etc. with their expanded forms.
- 4) Stop words list [12] was used to remove all non-informative stop words.
- 5) Emotions identifier as wow, awww, xxx (“many kisses”) or kkkkk (giggling) and laughter as hahaha, hehehe, jajaja and ahahaha also were replaced with their sentiment polarity.
- 6) All tweets were lowercased.
- 7) All digits and unnecessary punctuation were removed.
- 8) Repeated letters as yeeeees, yahooooo, loooooove were also removed.
- 9) We ignored all Non-ASCII characters.
- 10) All doubled tweets and retweets were removed.
- 11) Removed names of all businesses/companies according to the top brands on Twitter [13]. It turns out that when a company has many negative tweets about their customer service, the “probability” that any future tweet mentioning the same name is negative becomes huge.

### 3.3. Opinion Sentiment Classification

The most important step in this research is the selection of classifier for the text classification task. According to the paper [14], where author compares support vector machine (SVM) and multinomial Naïve Bayes (MNB) for both blog and micro-blog sentiment analysis, he finds that SVM outperforms MNB on blogs with long text but MNB outperforms SVM on micro-blogs with short text. Inspired by those results, in our research we selected Naïve Bayes method as our sentiment classifier.

The Naïve Bayes method for classification is often used in text classification due to its speed and simplicity. It makes the assumption that words are generated independently of word position. The Naive Bayesian classifier is a probabilistic model which is used for our purposes to estimate the probability that a tweet belongs to a specific class (positive, negative, or neutral). For a given set of classes, it estimates the probability of a class  $c^*$  given a document  $d$ , with terms  $t$ , as in Equation (1):

$$c^* = \arg \max_c P(c | d), \tag{1}$$

$$P(c | d) = \frac{P(c)}{P(d)} \prod_k P(t_k | c)$$

The parameters  $P(c)$  and  $P(t_k | c)$  are obtained through maximum likelihood estimates (MLE). The classifier then returns the class with the highest probability given the document.

### 3.4. Training

In this work for tweets collection, Twitter API was used. API has a parameter that specifies which language to

**Table 1.** Example of emoticons to be replaced using emoticon dictionary [11].

Icon	Meaning
:-) :) :o) :] :3 :c) :>=] 8) =) :} :^) :~)	Smiley or happy face.
:-D :D 8-D 8D x-D xD X-D XD =-D =D --3 =3 B^D	Laughing, big grin, laugh with glasses
:~)	Very happy or double chin
>:[ :-( :(-c :c :< :> C < :-[ :[ :{	Frown, sad
:-   :@ >:(	Angry
:'-( :(	Crying

retrieve tweets in. We had always set this parameter to English. Thus, our classification will only work on tweets in English because our training data is in English only. Throughout the course of this project about five million tweets were collected automatically to be used as training data.

In this study, to collect data for each class (“negative” and “others”, as for “others” class we use positive + neutral/objective tweets), positive “:)” and negative “:(” emoticons were used. There are multiple emoticons that can express positive and negative emotions. In the Twitter API, the query “:)” will return tweets that contain positive emotions and the query “:(” will return tweets with negative emotions. For the neutral training data set, we queried API with “http//” and “#hashtag”, because according to our own research almost all neutral/spam messages contain URL link and hash tags.

Tweets in our training set are from the time period from October to December, 2012. After the pre-processing step, we take the first 300,000 positive/neutral tweets (neutral tweets with neutral or spam content) and 300,000 tweets with negative content, for a total of 600,000 training tweets. On the basis of the extracted training data, we generate our sentiment classifier. We applied the Naïve Bayes algorithm to the classifier.

The challenging task of this research is that, sometimes users can express mixed sentiments in tweets toward product or services. For example, “Love iphone’s new design, but hate its short battery life ☹”.

Naïve Bayes classifier is useful for such cases, since it estimates probability of occurrences of each word in tweet. Thus, to not distort the initial meaning of tweet we do not remove slang and other informal language forms as in previous researches. For instance, the above mentioned tweet will look as following after all necessary pre-processing steps: “love new design hate short battery life [sad]”.

### 3.4. Testing

The test data was also collected automatically using the Twitter Search API. All set of the test data was manually marked as “others” or “negative”. Not all the test data has emoticons. We used the following process to collect test data.

- We searched the Twitter API with specific queries. These queries are arbitrarily chosen from different domains. For example, these queries consist of consumer products, services, and people. The query terms we used are listed in [Table 2](#).
- We looked at the result set for a query. If we saw a result that contains a sentiment, we mark it as “others” (positive/neutral) or “negative”. Thus, this test set is selected independently of the presence of emoticons.

## 4. Experiment and Results

Our experiment was conducted by gathering large amount of tweets using Twitter Stream API (from October to December, 2012), to be used as training and testing data. For the training set, data were collected by querying Twitter API for two types of emoticons:

- Smiley emoticon
- Frowny/Sad emoticon

Also, emoticon corpus from the work [6] was used additionally to our training set.

For the neutral dataset, objective tweets with no sentiment or tweets with spam context were considered as

**Table 2.** Query terms for the test data.

Product/Service	Tweets #
Air Asia	196 (negative: 47, others: 149)
Windows 8	168 (negative: 26, others: 142)
PSY	123 (negative: 12, others: 111)
Galaxy S III	168 (negative: 18, others: 150)
iPhone 5	210 (negative: 27, others: 183)
WiiU	146 (negative: 17, others: 129)

neutral. The collected dataset was used to extract features, which will be used to train our sentiment classifier. The product reputation was estimated by analysing the output result of classifier within given product name. For test data, tweets mentioning service, mobile phones, video game console, OS and popular music was used (**Table 2**). As in the paper [7], we adopt accuracy and F1-score as our evaluation metrics. Accuracy is a measure of what percentage of test data are correctly predicted, and F1-score is computed by combining precision and recall. The results of the evaluation are shown in **Table 3**. As you can see from the results, classifier’s accuracy trained using unigrams as features are quite high, but in case with negative set the F1-score is pretty low. To get more high accuracy for negative data set, we have tried to include bigrams (two-word combination) as classification features, not just unigrams (single word). Using bi-grams is supposed to help with tweets that contain negated phrases like “not good” or “not bad”. In our experiment, negation as an explicit feature with unigrams did not improve accuracy, so we are very motivated to try bi-grams. Below you can find the result’s comparison of F1-score for “Negative” (**Figure 1**) and “Others” (**Figure 2**) classification task for unigrams and bigrams classification features. From the results, we can say that including bigrams as classification features did not give us any good improvement in results. Bigrams tend to be very sparse and the overall accuracy drops in the case with Naïve Bayes classifier. In general, using bigrams as features is not useful because the feature space is very sparse. In the paper [6], the authors are also got unsatisfied results by using bigrams.

To improve our classifier’s result, we decided to build and use our own dictionary of negation phrases with its sentiment meaning. So the further step as building dictionary, with negation word as “not” and preceding adjectives to change its sentiment polarity, for example, “not bad”—“good”, “not annoyed”—“pleased” and etc. was included to the pre-processing steps.

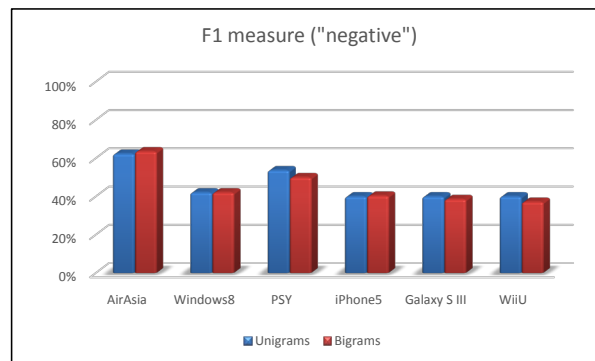
**Table 4** is the result we got by using negation phrases dictionary with unigrams as classification features. Further graphs are the comparison of results using three methods (**Figure 3** and **Figure 4**). From the results we can say that our proposed method with more concentration on the pre-processing step, as using emoticons, acronyms and etc. is practicable, especially in case with objective tweets during its difficulty in classification.

**Table 3.** Classifier accuracy and F-score for two way classification task.

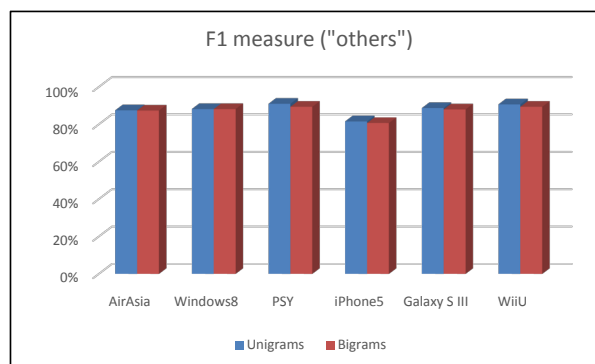
Product/Service	Accuracy	F1 measure	
		Other (Pos and Objective)	Negative
Air Asia	81.3%	87.5%	62.3%
Windows 8	80.5%	88.2%	42.1%
PSY	85.0%	91.0%	53.7%
Galaxy S III	71.8%	81.6%	40.0%
iPhone 5	81.0%	88.7%	40.0%
WiiU	84.0%	90.7%	40.0%

**Table 4.** The results of using negation dictionary.

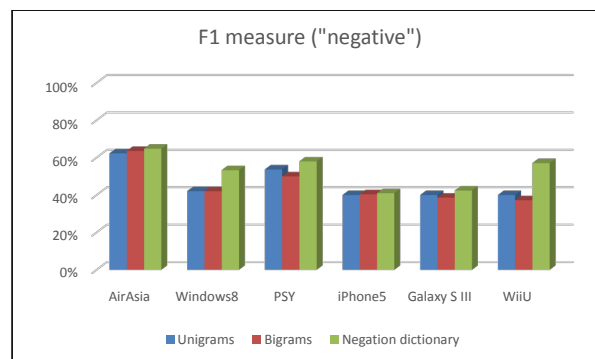
Product/Service	Accuracy	F1 measure	
		Other (Pos and Objective)	Negative
Air Asia	82.1%	88.0%	64.8%
Windows 8	83.3%	89.8%	53.3%
PSY	87.0%	92.3%	58.0%
Galaxy S III	72.6%	82.2%	41.0%
iPhone 5	72.8%	82.2%	42.4%
WiiU	87.7%	92.8%	57.1%



**Figure 1.** F1-score (“negative”) for unigrams and bigrams classification features.



**Figure 2.** F1-score (“others”) for unigrams and bigrams classification features.



**Figure 3.** F1-score results (“negative”) using three methods.

Sentiment classification toward product is the challenging one. Let’s have a look at tweet, mentioning iPhone 5: “I have to admit I’m a little jealous of robbies iphone 5 :-(-”. In general, it is negative tweet, but from the point of Apple Inc., it is positive tweet which tells, that their product is highly demanded.

## 5. Conclusions

Micro-blogging nowadays became one of the major types of the communication. A recent research has identified it as online word-of-mouth branding. The large amount on information contained in micro-blogging websites makes them an attractive source of data for opinion mining and sentiment analysis.

This study investigates how product reputation can be automatically extracted from famous Twitter micro-blogging service. We have proposed an approach based on opinion sentiment classification. We used the

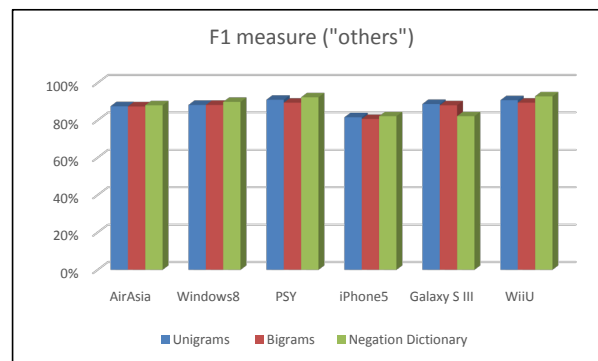


Figure 4. F1-score results ("others") using three methods.

collected corpus to train our sentiment classifier. Our classifier should be able to determine positive, negative and neutral sentiment from tweets and estimate the reputation of given product for the certain period of time.

As for the future work, we plan to collect data with detection of fake twitter accounts, to prevent fake reputation of product/services and make improvements in our approach to get high reputation accuracy.

## References

- [1] Pak, A. and Paroubek, P. (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, 1320-1326.
- [2] Pang, B., Lee, L. and Vaithyanathan, S. (2002) Thumbs up? Sentiment Classification Using Machine Learning Techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, 79-86.
- [3] Hu, M. and Liu, B. (2004) Mining and Summarizing Customer Reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177.
- [4] Wilson, T., Wiebe, J. and Hoffmann, P. (2005) Recognizing Contextual Polarity in Phrase Level Sentiment Analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347-354.
- [5] Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009) Twitter Power: Tweets as Electronic Word of Mouth. *American Society for Information Science and Technology*, **60**, 2169-2188.
- [6] Go, A. (2009) Twitter Sentiment Classification Using Distant Supervision. Technical Report.
- [7] Liu, K.-L., Ki, W.-J. and Guo, M. (2012) Emoticon Smoothed Language Models for Twitter Sentiment Analysis. *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 1678-1684.
- [8] Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R. (2011) Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Languages in Social Media*, 30-38.
- [9] Twitter Developers. <http://dev.twitter.com>
- [10] Personality Adjectives. <http://ja.scribd.com/doc/48820833/Personality-Adjectives>
- [11] List of Emoticons. [http://en.wikipedia.org/wiki/List\\_of\\_emoticons](http://en.wikipedia.org/wiki/List_of_emoticons)
- [12] List of English Stop Words. <http://norm.al/2009/04/14/list-of-english-stop-words/>
- [13] Minato, C. (2012) The Top 20 Brands on Twitter, Ranked by Influence with Customers. <http://www.businessinsider.com/the-top-20-brands-on-twitter-ranked-by-influence-with-customers-2012-7?op=1>
- [14] Birmingham, A. and Smeaton, A.F. (2010) Classifying Sentiment in Microblogs: Is Brevity an Advantage? *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 1833-1836.



Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).

