

# Review on Brand Related Sentiment Analysis in Commercial Field

Digamber Shelke, Sachin Deshmukh

M.Tech, Department of Computer Science and Information Technology,  
Dr. BAMU, Aurangabad 431004  
[digambershelke4@gmail.com](mailto:digambershelke4@gmail.com)

**Abstract:** We present an approach to brand-related Titan review as dataset sentiment analysis using feature. The approach add associated with the unique characteristics of the R language, and the recall mild sentiment expressions that are of interest to brand management practitioners. We demonstrate the effectiveness of the approach on an Titan brand-related review dataset. to find trend in market. to find sentiment of brand related .and to find likes and dislike of ppl sentiment. Solution for this we have used PMI alorithem and SVM classifier which is giving accuracy of 68% accuracy.

**Keywords:** Feature engineering, n-gram analysis, Machine learning, SVM, Titan

## Introduction:

Sentiment analysis, also called *opinion mining*, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity

analysis, affect analysis, emotion analysis, review mining, etc. However, they are now all under the umbrella of sentiment analysis or opinion mining. While in industry, the term *sentiment analysis* is more commonly used, but in academia both *sentiment analysis* and *opinion mining*.

We have gather Titan brand review dataset. i hv applying PMI, and SVM aglorithem review classify into positive negative and neutral .then feature will identified using feature engineering or other feature techniques. Action taken on preprocessing task of I have remove comma, "", <span>, hash tag, space, missing column. Ppl buy according to brand ,their quality .here we have to see which brand ppl follow and quality of brand .we analyze emoticon of ppl, they will buy or not product of industry. it helpful for industry in commercial point of view.

**Related Work:** [1] Ghais has done brand sentiment analysis. he has used feature engineering techniques for finding sentiment. [2] I am using svm techniques for accuracy [3] Pmi algorithem are used as classifier to classify positive and negative class. [4] Sentiment Analysis in Twitter. System adopts a hybrid classification process that uses three classification approaches: rulebased, lexicon-based and machine learning approaches. They suggest a pipeline architecture that extracts the best characteristics from each classifier.

**Methodology :** We have gather Titan tweet review dataset. i applying applying naive bias aglorithem review classify into positive negative and neutral .then feature will identified using feature engineering or other feature techniques. Action: preprocessing task of I have remove comma, "", <span>, hash tag, space, missing column .feature will be done. then postaging will apply.

## III. DATA COLLECTION, PREPROCESSING & METHODOLOGY

### A. Data Collection and Preprocessing

Data set [8] consist of 1009 text document of movie review. In data pre-processing, all text was converted to lowercase, for making the data same. Then meaningless words are converted to meaningful words wherever possible. e.g. good converted to good, unwanted punctuations such as comma, numbers i.e. un-necessary data should be removed. After this we have to perform POS tagging. POS tagging is nothing but the part of speech tagging in which we tag each word to get its part of speech. It will help us to select particular word of particular part of speech. After pre-processing work, feature extraction process is carried out. In feature extraction we have to extract

the phrases of particular pattern appear in the sentence/text document .The pattern is as follows which is used by Turney in 2002 [7]. We used POS tagging method as earlier discussed.

**B. Methodology for Sentiment Analysis**

**1. SO-PMI based approach:**

This method calculates the PMI i.e. Point-wise Mutual Information between two words and produce numeric score. The formula is as follows.

$$PMI(word1,word2) = \log_2(\text{prob}(\text{word1}\&\text{word2})/\text{prob}(\text{word1})*\text{prob}(\text{word2})) \dots\dots\dots (1)$$

Here, prob(word1 & word2) is the probability of word1 and word2 co-occur in the sentence/text document.

Scoring for semantic orientation:

Here, PMI(word1,positive word) and PMI(word1,negative word) calculated, so that we can calculate semantic orientation score.

$$SO\ Score = PMI(\text{word1, positive word}) - PMI(\text{word1, negative word}) \dots\dots\dots (2)$$

All phrase value SO Score is calculated and by averaging it we can get average numeric score ,if that is

**2. Word Based Approach:**

Apply stemming procedure on text data for converting derived words to their root by removing end characters and then calculate the polarity of review by comparing the positive and negative word list [9]. If count of positive words is more than negative word in review then review is positive, if less then review is negative otherwise neutral.

**IV. EXPERIMENTAL ANALYSIS**

We have used movie review dataset of 1009 review. In SVM Cross validation, training data and testing data is of 70:30 pattern i.e. 700 review for training data and remaining 409 for testing data. The accuracy value represents the percentage of test texts which were classified correctly by the method. Twitter model gives 66% to 68% accuracy value for titan tweets .Here we have taken different of model for analysis.Table-2 shows different model’saccuracy preprocessing is done for removing the unwanted data to minimize the data size. After preprocessing step, selection of phrase value carried out according to pattern discussed in Tab

**Model**

Model	Svm Cross validation		
	3fold	5 fold	10fold
SO-PMI	69.20	69.32	71.72
Word Based Approach	66.52	70.35	69.52
Hybrid(SO-PMI and Word Based Approach)	70.91	70.12	70.80
Hybrid SVM(SO-PMI, Word Based Approach and SVM)	67.29	70.61	71.9

**Word Based Approach** 66.52 70.35 69.

**Conclusion:**This paper makes several contributions to Twitter sentiment analysis, demonstrated through application on a corpus of review related to the amazon brand. Earlier research on Twitter classification classified factual sounding tweets as a neutral review. Using this approach, they state that “more than 80%” of his review contain no sentiment. Our approach to sentiment analysis has

increased sensitivity, accounting for review with mild sentiment (positive and negative), resulting in a more accurate identification of the neutral category.

## REFERENCES:

- [1] Determining stock prediction using is sentiment analysis by Mittal
- [2] NILC USP: A Hybrid System for Sentiment Analysis in Twitter Messages Pedro P. Balage Filho and Thiago A. S. Pardo Interinstitutional Center for Computational Linguistics (NILC)
- [3] The Role of Media in the Stock Market Professor Paul Teetlock News and Finance Conference March 2016
- [4] DATA TREATMENT WHITE PAPER Preparing data for analysis using R Nina Zumel, Win-Vector LLC March 2016 Win
- [1] Twitter, Inc. (2013). IPO Prospectus. Downloaded on February 2014 (<http://www.sec.gov/Archives/edgar/data/1418091/000119312513390321/d564001ds1.htm>).
- [2] Jansen, B., Zhang, M., Sobel, K., and Chowdury, A. (2009). Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*, 60.
- [3] Abbasi, A., Hassan, A., and Dhar, M. (2014). Benchmarking Twitter Sentiment Analysis Tools. *Proc of LREC Conf.*
- [4] Ghiassi, M., Skinner, J., and Zimbra, D. (2013). Twitter Brand Sentiment Analysis: A Hybrid System using Ngram Analysis and Dynamic Artificial Neural Network. *Expert Systems with Applications*, 40.
- [5] Ghiassi, M. and Saidane, H. (2005). A Dynamic Architecture for Artificial Neural Network. *Neurocomputing*, 63.
- [6] Sentiment140 (2015). [www.sentiment140.com](http://www.sentiment140.com).
- [7] Go, A., Bhayani, R., and Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. Technical Report, Stanford Digital Library Technologies Project.
- [8] Repustate (2015). [www.repustate.com](http://www.repustate.com).
- [9] Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62.
- [10] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs Up?: Sentiment Classification using Machine Learning Techniques. *Proc of the ACL Conf on EMNLP*.
- [11] Gamon, M. (2004). Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis. *Proc of Conf on Computational Linguistics*.
- [12] Abbasi, A., Chen, H., and Salem, A. (2008). Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. *ACM Transactions on Information Systems*, 26(3).
- [13] Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proc of ACL Conf*.
- [14] Kim, S. and Hovy, E. (2004). Determining the Sentiment of Opinions. *Proc of the Intl Conf on Computational Linguistics*.
- [15] Pak, A. and Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proc of LREC Conf*.
- [16] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. (2011). Sentiment Analysis of Twitter Data. *Proc of ACL HLT Conf*.
- [17] Mittal, A. and Goel, A. (2012). Stock Prediction Using Twitter Sentiment Analysis. Working Paper, Stanford University.
- [18] Barbosa, L. and Feng, J. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. *Proc of COLING Conf*.
- [19] Kouloumpis, E., Wilson, T., and Moore, J. (2011) Twitter Sentiment Analysis: The Good the Bad and the OMG! *Proc of AACL Conf on Weblogs and Social Media*.
- [20] Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L., and Hughes, M. (2013). Sentiment Analysis of Political Tweets: Towards an Accurate Classifier. *Proc of ACL Workshop on Language in Social Media*.
- [21] Hu, Y., Wang, F., and Kambhampati, S. (2013). Listening to the Crowd: Automated Analysis of Events via Aggregated Twitter Sentiment. *Proc of Conf on Artificial Intelligence*.