INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Sentimental Analysis of Tweets using Principle Component Analysis Technique

Dr. Mohd. Abdul Raffey[1], Mr. Pankaj N. Patil[2], Dr. Sunil Kawale[3]

[1, 2, 3]*Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India*

Abstract: The much commonly famous and well-known micro-blogger web platform is twitter where registered user over the world can share their opinions by twitting. Users sharing the reviews on products which they are using and sometimes they exaggerate tweets about the same product. Sentiment Analysis is the technique of identifying the validity of such tweets by analyzing hidden emotions in their tweets Sentiment Analysis is widely used to analyze emotions in Text, News, Figures and Pictures etc…
Keywords: Sentiment Analysis, feature selection methods, machine learning, twitter, reviews, PCA

I. INTRODUCTION

Opinion Mining is currently the ongoing field of study. Many researchers are working on it for analyzing emotions, sentiments or views. Another challenging field attracting to researchers is that sentiment analysis, It is a process of decides the statement or a piece of statement is positive, negative or neutral. Nowadays this technique is widely used to discover what people feel about the particular product. Various industries are working on Sentiment Analysis to improve their sales by analyzing the emotions of users by text, news, figures and pictures.

A. The process of sentiment analysis is performed at different levels such as
1) Document Level: Documents expressed opinions like web reviews or open-ended questions in the market survey; the main task is to find whether the whole document is positive, Negative or Neutral.
2) Sentence Level: This level the sentences are checked for its Positive, Negative or Neural biased.
3) Aspect Level: At this level aspect driven sentiment analysis identifies the aspect of a given target entity and estimate the sentiment polarity for each mentioned aspect.

The most researchers work with Sentiment Analysis technique carried on news and text. The sentiment analysis on Twitter is a challenging work. Twitter is the very famous micro-blog website using by many people around the world. Twitter is founded in 2006 by Jack Dorsey, Noah Glass, Biz Stone and Evan William at SanFrancisco, California USA.

Principle Component Analysis (PCA) is a dimensionality reduction technique that is commonly used to transform a high-dimensional sample to small dimension by finding a new set of variables, smaller than the original variable without changing the meaning of sample.

B. The PCA used before running on Machine Learning Algorithms on the data because
1) It reduces the size of space on which K-nearest-neighbors (KNN) must calculate the distance it helps to performance improvement.
2) It also reduces the number of degree of freedom of hypothesis which reduces the risk of overfitting.
3) If the data/sample is in less amount most algorithms are faster.
4) When you reduce using PCA it simplifies the data set.

II. LITERATURE REVIEW

Bo Pang, at.al. (2002) used Machine Learning, Naïve Byes and Space Vector Machine (SVM) for Sentiment Analysis they found SVM is the best method and Naïve Byes is worst. Jeonghee Yi, Tetsaya Nasukwa (2003) presented Sentiment Analyzer through Neural Language Processing (NLP) technique and shows an accuracy of 87% for review, 86% to 91% for general web pages and 91% to 93% for news articles. Jeevanandam Jotheeswaran, Loganathan R. and Madhu Sudhanan B. (2012) reduced the data set through PCA and shows the effectiveness of preprocessing of movie review LVQ 75% accurate which is better than Naïve Byes. Akshi Kumar and Teeja Mary Sebastian (2012) purposed paradigm to investigate on twitter data, used a corpus-based method to adjectives and dictionary-based method on verbs and adverbs. Anju Tivari (2015) used PCA algorithm to reduce the textual reviews

(BOW and BON) and SVM classifier used to prove much accuracy. G. Vinodhini, R.M. Chandrasekaran (2015) has done the hybrid combination of PNN and PCA for reducing the training time and increases the classification performance, ANN is better in the training time. Vikram Kumar (2016) evaluated Naïve Byes of Linear Vector Quantization (LVQ) of the reducing data set by PCA. Bholane Savita (2016) used two methods senti-strength and SVM to find the accuracy of interpretation and proven that SVM is better. Neha Upadhyay, Prof. Angad Singh (2016) used Naïve Byes, SVM, Maximum Entropy and Ensemble Classifier to estimate the probability for better result taken electronic items for Positive, Negative and neutral classification to help peoples for selecting the best item. G. Vinodini (2017) has taken a review of the work on PNN and PCA.

A. Proposed Model

In this model, we have taken 2000 tweets regarding 10 features of iPhone 7. These tweets have been selected randomly which are recorded on twitter microblogger website and categorized into 6 different types of emotions proposed by Ekman's which are given in table no. 1.

	Happy	Disgust	Surprised	Sad	Angry	Fear	Natural
1	318	248	125	365	475	238	231
2	265	752	136	125	251	240	231
3	198	42	148	147	356	847	262
4	175	354	179	185	258	634	215
5	642	258	168	673	103	4	152
6	371	456	152	259	150	356	256
7	586	369	475	246	160	40	124
8	476	147	283	864	14	43	173
9	412	546	412	245	148	53	184
10	222	652	751	125	125	110	15

Table 1: Data Set

In general, this information is vague to decide the real approach of users about iPhone 7. Hence here we want to find real reviews about products of users for that we proposed Principle Component Analysis (PCA) to find the association of different emotions, to decide which motions are closely associated. PCA has been performed through SPSS.

Correlation matrix (Table No. 2) shows the association between emotions. Here the showing negative relation found in "Happy" & "Disgust" also "Happy" & "Sad"

	Happy	Disgust	Surprised	Sad	Angry	Fear	Neutral
Happy	1.000	.028	.045	.052	-.192	-.720	-.226
Disgust	.028	1.000	-.401	-.510	-.181	-.346	-.330
Surprised	.045	-.401	1.000	-.193	-.443	-.448	-.005
Sad	.052	-.510	-.193	1.000	-.475	-.475	-.054
Angry	-.192	-.181	-.443	-.475	1.000	.578	.510
Fear	-.720	-.346	-.448	-.475	.578	1.000	.587
Neutral	-.226	-.330	-.005	-.054	.510	.587	1.000

Table 2: Correlation Matrix

Communalities Table No. 3 that it shows all values of variables is greater than 0.08 i.e. Happiness (80%) and Surprised (83%) are the highest value of communalities means these are significant meaning compared to other variables.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

	Initial	Extraction
Happy	1.000	.802
Liked	1.000	.626
Surprised	1.000	.831
Sad	1.000	.886
Angry	1.000	.670
Loves	1.000	.807
Natural	1.000	.818
Extraction Method: Principal Component Analysis.		

Table 3: Communalities

In Table No. 4 shows that the extraction sums of square loadings contain only that Eigenvalue which is greater than 1, percentage of the variance of factors shown in the same table which is 46.75% and 30.97% respectively.

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.273	46.755	46.755	3.273	46.755	46.755	2.773	39.621	39.621
2	2.168	30.976	77.731	2.168	30.976	77.731	2.668	38.110	77.731
3	.760	10.859	88.590						
4	.468	6.679	95.269						
5	.260	3.712	98.981						
6	.071	1.019	100.000						
7	9.194E-17	1.313E-15	100.000						
Extraction Method: Principal Component Analysis.									

Table 4: Total Variance Explained

Factors determined through PCA are also checked by Scree Plot which shows the two components are considered in a determination of PCA. Hence the PCA used in this analysis is very much correct and PCA is used to reduce the big data to the small groups of factors.
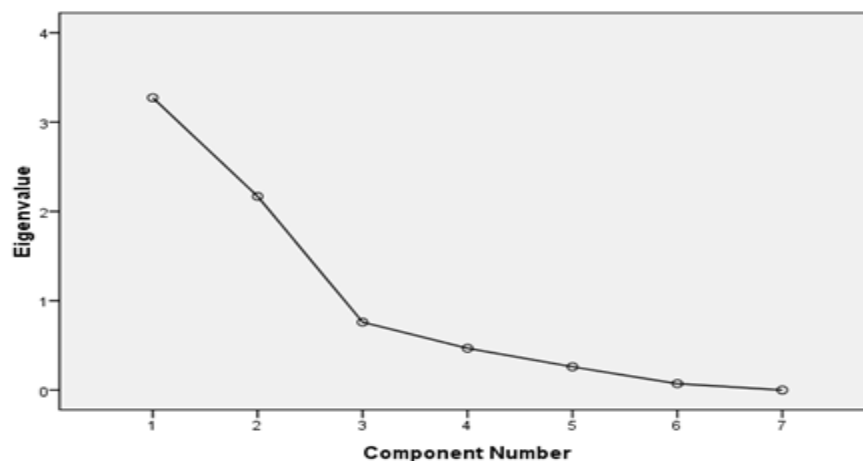

Fig. Screen Plot

III.   CONCLUSION

PCA is the very useful technique for data reduction in Sentimental Analysis in view to deciding the real approach of reviews PCA is applied on 2000 tweets related to iPhone 7 to find the reliability of tweets. In this paper we applied PCA technique to find out sentiments in tweets the eigenvalue shown in table 4 is help us to take decision about the positive, negative or neutral tweets of iPhone reviews so that new buyer can take decision with the help of table 4 and can decide whether he has to buy iPhone or not. The accuracy is 77% of reviews over 2000 tweet that means many people gives their feedback through tweets in the real sense.

REFERENCES

[1]   Bholane Savita D., Prof. Deipali Gore , "Sentiment Analysis on Twitter Data Using Support Vector Machine" in I JCST 2016.

[2]   G. Vinodhini, R.M. Chandrasekaran "A comparative performance evaluation of neural network based approach for sentiment classification of online reviews" in JKSU–C&IS, pp. 2–12 2016.

[3]   Bo Pang and Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques" in EMNLP, pp. 79-86 July 2002.

[4]   Yi, J., Nasukawa, T., Niblack,W., & Bunescu, R., "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques", in ICDM, pp. 427–434, 2003.

[5]   Kirutika M., sanjana woonna, Priyanka Giri, "Sentiment analysis on twitter data" in IJIET 2016, pp. .

[6]   Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey" in ASEJ 5, 1093–1113, 2014

[7]   Akshi Kumar and Teeja Mary Sebastian, "Sentiment Analysis on Twitter" in IJCSI pp. 372-378, 2012.

[8]   Neha Upadhyay, Prof. Angad Singh, "Sentiment Analysis on Twitter by using Machine Learning Technique" in IJRASET

[9]   Ekman, P.: An Argument for Basic Emotions. Cognition and Emotion. 6, 169–200 (1992)

[10] A Abbasi, HC Chen and A Salem. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. ACM Transactions On Information Systems, Volume 26, Number 3, 2008.

[11] Tan, S. B., & Zhang, J. (2008). An Empirical study of opini on analysis for Chinese documents. Expert Systems with Application, 34(4), 2622–2629.

[12] Hu, M. & Lui, B. (2004), "Mining and Summarizing Customer Reviews". In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004), p. 168–177.

[13] Haseena Rahmath P, "Opinion Mining and Sentiment Analysis -Challenges and Applications" in IJAIEM pp. 401-403, 2014.

[14] https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/

[15] https://klevas.mif.vu.lt/~tomukas/Knygos/principal_components.pdf

[16] https://www.cs.cmu.edu/~elaw/papers/pca.pdf

[17] http://support.sas.com/publishing/pubcat/chaps/55129.pdf

[18] http://www.lauradhamilton.com/introduction-to-principal-component-analysis-pca

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089   (24*7 Support on Whatsapp)