

Analysis of Products Customized Features

Jahnvi Gor¹, Prof. Prashant B. Swadas², Assi. Prof. Kanu G. Patel³

¹PG Student, BVM, V.V. Nagar, Gujarat, India.

²Computer Department, BVM, V.V. Nagar, Gujarat, India.

³IT Department, BVM, V.V. Nagar, Gujarat, India.

Abstract - Opinion mining it's like game of words / corpus. Different persons are handling this with their own perspective model with the help of classic technique. Here we could find some of the techniques have been seen during survey.

Keywords- Feature extraction, classification, Opinion mining, svm, k-NN.

I INTRODUCTION

The market growth of internet attracts many types of research opportunities like data mining, distributed system, etc. Among them data mining has one interesting concept that is opinion mining.

Opinion mining has mostly done at document, sentence, and corpus, feature level using various classifiers & clustering techniques. Survey has been focused on different techniques & related to improvement of lazy learning classifier. Various perspectives are used here like linguistic analysis, comparative study, keyword selection, feature extraction etc.

II CLASSIFICATION ALGORITHMS

Mainly two types are [22]: Supervised and Unsupervised classification algorithms.

Among them some commonly used classifiers are:

1 K-NEAREST NEIGHBOR ALGORITHM: [23]

Nearest neighbor Algorithm or instance based algorithm are used to find the nearest neighbor & based on this it will calculate the distance & make a class of it. As shown in figure white circle is unknown instance & red-Negative & Blue-Positive.

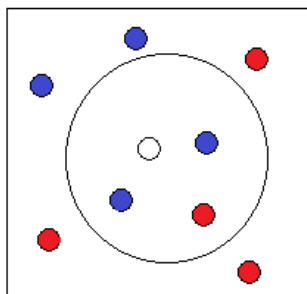


Figure-1: Shows the nearest neighbor

Here circle is some criteria which is predefined & red dot shows the negative point so based on this we could

see that blue is more in circle then red so unknown set are goes with blue circle class, likewise nearest neighbor algorithm work. Nearest Neighbor algorithm has various types like Modified k-NN & reduced k-NN & Weighted k-NN etc. Many of them are used for Image processing purpose & in Image mining also.

In k-NN with binary method, for large feature vector Euclidian distance doesn't work better.

2. Support Vector Machine (SVM): [1]

Given a category set, $C = \{+1, -1\}$ and two pre-classified training sets, i.e [1], a positive sample set, $Tr+ = \sum_{i=1}^{n_1} (d_i, +1)$ and a negative sample set, $Tr- = \sum_{i=1}^{n_2} (d_i, -1)$, the SVM finds a hyperplane that separates the two sets with maximum margin (or the largest possible distance from both sets). At pre-processing step, each training sample is converted into a real vector, x_i that consists of a set of significant features representing the associated document, d_i . Hence,

$TR+ = \sum_{i=1}^{n_1} (x_i, +1)$ for the positive sample set and $TR- = \sum_{i=1}^{n_2} (x_i, -1)$ for the negative sample set. In this regard, for $c_i = +1$, $w \cdot x_i + b > 0$, and for $c_i = -1$, $w \cdot x_i + b < 0$.

Hence, $T = \{c_i \cdot (w \cdot x_i + b) - 1\}$ becomes an optimization problem defined as follows: minimize $(1/2) \|w\|^2$, subject to $c_i \cdot (w \cdot x_i + b) - 1$. The result is a hyper plane [1] that has the largest distance to x_i from both sides. The classification task can then be formulated as discovering which side of the hyperplane a test sample falls into. From this concept [20], here they analyze the feature of product with cosine similarity function & they use this for set margin for svm select the appropriate result to show that which feature has how many positive review & how many negative review. Here, lemmatization with svm classifier is seen during survey but this concept of finding word root with [9] modified k-nearest neighbour classifier was not used to obtain the word root. So proposed is at focus.

Other techniques are also used like association rule mining [22] etc. The general process shown in figure below,

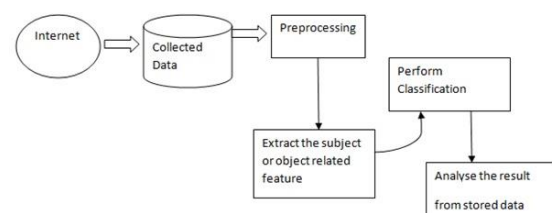


Figure-2:General Process

Here, after collecting the data from web or document or review of any product, news articles, blogs, forums etc are then pre-processed according to usage. Unnecessary alphabets are removed so parsing will be easy & than useful information related to feature level or direct & indirect level of sentence & word are analyzed to develop well trained data from this data will feed to classifier to improve accuracy. This is the general concept of opinion mining. Also author are focused now a days at building the synonymous of feature to improve classifier result also some interest in improvement towards lazy classifier have been seen so based on this proposed focus is at this direction.

Level [2]:

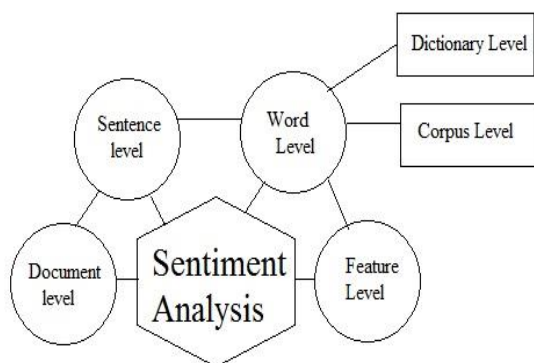


Figure-3: Level of Sentiment Analysis

III LITERATURE REVIEW

In opinion mining related to our work Survey shows present methods,

1 Ankita Srivastava, Dr. M.P. Singh, Prabhat Kumar [9], Python & NLTK tool kit is used to perform the computation over the dataset of mobile product. After collecting the data from different site like amazon.com etc. pre-processing are performed to remove stop word, by thoroughly analyze the dictionary & create weight for each adjective ,noun ,verb, past tense , they create positive, negative, neutral data file & then in first parse after creating train set by analyzing positive & negative dataset & assign positivity score to each sentence, according to predefined margin they generate positive & negative data file, this data files are again analyzed with neutral dataset to cover high, mid, weak polar reviews & generate the sentence score. But while removing stop word explicitly negation are less considered negation are of two types content & functional, content are handled But functional are not, which might change polarity. By considering both type of negation content (i.e. worst) & functional (i.e. is not worst) result would be analyzed. Most important thing they give percentage score at previously to the whole review for once but proposed will be analyzed features; according to result will be re-ranked the sentences.

2. Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Prashant Kumar Singh [21] To analyze the big amount of data author analyze the review with different approach by analyzing the data line by line & when they find sentiment word they calculate of average word weight & give score to feature & analyze that score with

Google analyzer meter interestingly & also a helpful approach for manufacturer of product .But polarity shifting could be added.

3. Taysir Hassan A. Soliman, Mostafa A. Elmasry, Abdel Rahman Hedar, M.M.Doss, [20], Here Author utilize the Support Vector Machine with different cosine Similarity function & also create the feature synonym dictionary to analyze the feature of product with svm easily. This was with subjective analysis which is on direct statement but focus would be moved with explicit statement & what reverse polarity affects!?. Following table shows some other survey.

AUTH-OR	DATA-SET	GRANU-LARITY LEVEL	KEY IDEA
Shoush-an Li, et. Al.[10] 2013	Product, movie, multi-domain Reviews	Corpus level	Bag-of-word model, polarity shifting, novel term counting based classifier
Michael Gamon, et.al. [16] 2005	Car review	Sentence level	Based on taxonomy it will extract sentence & makes cluster for visualization & classify in positive , negative, other class
Ankita Srivas-tava, et.al. [9] 2014	Product review	Sentence level	Improvement over k-nn ,lazy learning classifier
Jingye wang, Heng Ren[15] 2007	Product review	Word level	Experiment at Semantic level
Pooja Kherwa et. Al.[21] 2014	Products,laws, policies'; reviews, discuss-ion, forums etc.	Word level	Find the noun & adjective, calculate average score
Mrs. Vrushali Yogesh Karkare, Dr. Sunil R. Gupta[6] 2014	Product review	Corpus level	Based on feature rating , compare products

Tushar Ghorpade, Lata Raghya[8] 2012	Hotel review	Word level	They improve training set by analyzing word & then fed to NB classifier to improve accuracy
Prabhu palanisamy et.al.[14] 2013	Product & social issue	Sentence & word level	Verb & Adverb level
Ming-xing Wu , Liya Wang & Li Yi[11] 2013	Product review	Feature level	Analysis was done on unstructured data which give structured feature-opinion pair o/p.
M.A.Jawale et. Al.[13] 2014	Product review	Feature level, Sentence level	Automatic review classification

IV CONCLUSION

In Opinion mining Field one will analyze result [2] by calculating the precision, Recall, Accuracy & also comparing the result with other method to measure improvement. Now a day's authors are focusing on feature dictionary & also towards the improvement over classifier.

REFERENCES

- [1] S Padmaja, Prof. S Sameen Fatima."OPINION MINING AND SENTIMENT ANALYSIS –AN ASSESSMENT OF PEOPLES' BELIEF: A SURVEY", International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.4, No. 1, February 2013,10.5121/ijasuc.2013.4102.
- [2] Akshi Kumar, Teeja Mary Sebastia, "SENTIMENT ANALYSIS: A PERSPECTIVE ON ITS PAST PRESENT AND FUTURE", N.I. INTELLIGENT SYSTEMS AND APPLICATIONS, MECS,PAGE 1-14,2012.
- [3] Erik Cambria, Bjorn schuller, Yunqing Xia, Catherine Havasi, "NEW AVENUES IN OPINION MINING AND SENTIMENT ANALYSIS", MARCH/APRIL 2013, 1541-1672/13/\$31.00, Published by the IEEE Computer Society, , 2013 IEEE.
- [4] Nitin Bhatia, Vandana, "SURVEY of NEAREST NEIGHBOR TECHNIQUES", (IJCSIS) International Journal of Computer Science and INFORMATION SECURITY, Vol. 8, No. 2, 2010.
- [5] Ayesha Rashid, Naveed Anwer, Dr. Muddaser Iqbal, Dr. Muhammad Sher,"A SURVEY PAPER: AREAS, TECHNIQUES AND CHALLENGES OF OPINION MINING", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.
- [6] Mrs. Vrushali Yogesh Karkare, Dr.Sunil R. Gupta."PRODUCT EVALUATION USING MINING AND RATING OPINIONS OF PRODUCT FEATURES", 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies, IEEE.
- [7] Bing Liu,"Sentiment Analysis: A Multi-Faceted Problem", IEEE intelligent System, 2010.
- [8] Tushar Ghorpade, Lata Raghya."FEATURE BASED SENTIMENT CLASSIFICATION FOR HOTEL REVIEWS USING NLP AND BAYESIAN CLASSIFICATION", 2012 International Conference

- on Communication, Information & Computing Technology, Oct. 19-20, 2012 IEEE.
- [9] Ankita Srivastava,Dr. M.P. Singh, PrabhatKumar, "SUPERVISED SEMANTIC ANALYSIS OF PRODUCT REVIEWS USING WEIGHTED k-NN CLASSIFIER", IEEE 2014 11th international conference on information technology: new generations,978-1-4799-3187-3/14 \$31.00 © 2014 IEEE.
- [10] Shoushan Li,, Zhongqing Wang , Sophia Yat Mei Lee, Chu-Ren Huang, SENTIMENT CLASSIFICATION WITH POLARITY SHIFTING DETECTION, 2013 International Conference on Asian Language Processing, IEEE 2013.
- [11] Mingxing Wu and Liya Wang , Li Yi."A NOVEL APPROACH BASED ON REVIEW MINING FOR PRODUCT USABILITY ANALYSIS", 2013 IEEE.
- [12] Tanvir Ahmad, Mohammad Najmud Doja."OPINION MINING USING FREQUENT PATTERN GROWTH METHOD FROM UNSTRUCTURED TEXT", 2013 International Symposium on Computational and Business Intelligence, 2013 IEEE.
- [13]M.A.Jawale, Dr.D.N.Kyatanavar, A.B.Pawar, "IMPLEMENTATION OF AUTOMATED SENTIMENT DISCOVERY SYSTEM", IEEE international Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), May 09-11, 2014, Jaipur, India.
- [14] Prabu Palanisamy, Vineet Yadav, Harsha Elchuri."SERENDIO: SIMPLE AND PRACTICAL LEXICON BASED APPROACH TO SENTIMENT ANALYSIS".
- [15] Jingye Wang, Heng Ren."FEATURE-BASED CUSTOMER REVIEW MINING".
- [16] Michael Gamon, Anthony Aue, Simon Corston-Oliver, Eric Ringger."PULSE: MINING CUSTOMER OPINIONS FROM FREE TEXT", Natural Language Processing, Microsoft Research, Redmond, WA 98052, USA.
- [17]Wen Zhang, Taketoshi Yoshida, Xijin Tang."TEXT CLASSIFICATION BASED ON MULTI-WORD WITH SUPPORT VECTOR MACHINE", 2008 Elsevier.
- [18] RamachandraRao Kurada, Dr. K Karteeka Pavan, M Rajeswari,M Lakshmi Kamala."NOVEL TEXT CATEGORIZATION BY AMALGAMATION OF AUGMENTED k-NEAREST NEIGHBOURHOOD CLASSIFICATION AND k-MEDOIDS CLUSTERING" ,International Journal of Computational Science and Information Technology(IJCSITY) Vol. 1,No.4,November 2013.
- [19] Kunpeng Zhang, Ramanathan Narayanan, Alok Choudhary."VOICE OF THE CUSTOMERS: MINING ONLINE CUSTOMER REVIEWS FROM PRODUCT FEATURE-BASED RANKING",
- [20] Taysir Hassan A. Soliman,Mostafa A. Elmasry,Abdel Rahman Hedar,M.M.Doss,"UTILIZING SUPPORT VECTOR MACHINES IN MINING ONLINE CUSTOMER REVIEWS",ICCTA 2012,13-15 October 2012.
- [21] Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Prashant Kumar Singh."AN APPROACH TOWARDS COMPREHENSIVE SENTIMENTAL DATA ANALYSIS AND OPINION MINING", 2014 IEEE.
- [22] http://en.wikipedia.org/wiki/Machine_learning
- [23] http://wikipedia.org/wiki/K-nearest_neighbours_algorithm