International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 3 | Issue 1 | ISSN : 2456-3307

Intelligence Approaches for Sentimental Analysis in Social Networks : A Survey

Karthik R.V, S. Ganapathy
School of Computing Science and Engineering, VIT University-Chennai Campus, Chennai, India

ABSTRACT

Social network occupies an important place and also takes a considerable amount of time in people's daily lives. It has become so popular that people are sharing a huge amount of data and opinion on social network/review sites which in turn helps to find interesting insights for organizations/ vendors or consumers. In this Survey paper, we have presented the researches done in the area of Sentimental analysis / Opinion mining from Social network data. We also documented the outcome of comparative analysis, followed by our future work in Online product recommendations with help of Sentiment analysis.
Keywords : Social Network, Sentiment Analysis, User Reviews, Recommendation, Opinion.

I. INTRODUCTION

Social network and online shopping have become very popular these days. In Social networking sites, people share a lot of information in the form of images, opinions, ideas, etc. In many cases, they share their own experiences about a product or service and these online opinions are valuable as they play an important role in influencing a consumer's buying decisions especially in online shopping. A lot of research is currently happening to extract opinion insights. The collective term for this is known as Sentimental analysis. This paper showcases the importance of sentimental analysis in social networks and online shopping.

Research shows that people spend a large amount of time on social networking sites. According to the report by Common Sense Media, teens are spending more than 9 hours on an average in online.

Online shoppers usually look for reviews and feedbacks before buying a product. This review data is found in abundance in social networks. This information is highly valuable to online sellers and shopping websites as it gives an insight into their products' feedback and consumer expectations.

In social networks, it is often found that feedbacks are in unstructured form. Users often use annotation, hashtags, images, animated gifs and mixed languages, so it is important to handle all these formats that come from different social network sites through intelligent techniques. The techniques should be generic and handle a large volume of unstructured data efficiently. Sentiment analysis or opinion mining helps to make decisions based on the user opinions and reviews. There are different levels of Sentiment Analysis.

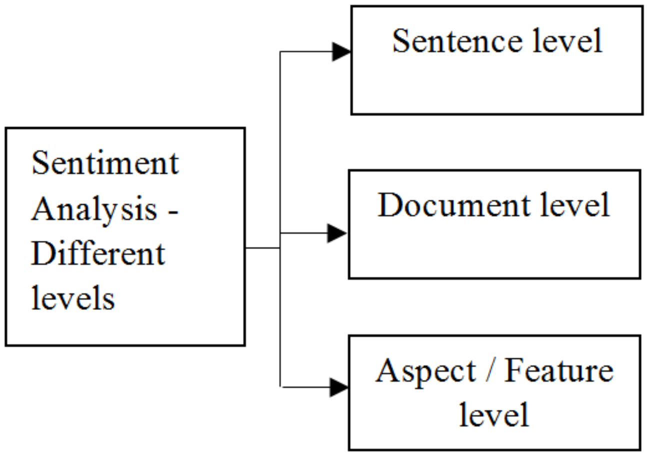

Figure 1 : Sentiment Analysis different levels

Feature selection is used to reduce the dimensionality of the vast data. It detects all the relevant features and discards the irrelevant ones. Feature Selection technique has several benefits: Improves the performance of machine learning, helps in data reduction, reduces computational burden and reduces storage requirement, and improves the accuracy of prediction. There are two approaches in feature selection: Filter based approach and Wrapper based approach.

In order to categorize the opinions as positive or negative, it is important to identify and classify the opinion/ sentimental words like 'important' 'bad', 'wonderful', 'awesome', 'horrible', 'terrible', 'pleasant', etc. it finds group of models, in which it could categorize and distinguish the data classes such that these models are used to predict the object for the unknown class labels.

Our proposal work here is to recommend feature product ranking based on customer's specific demographic information (age group, profession etc.,) and occasions (Festivals etc.,) through Sentiment Analysis.

II. LITERATURE SURVEY

1.1 Works on Social Network

Social networks are a huge repository of user opinions and feedbacks. Availability of such huge volume of data provides an opportunity to explore and extract useful information which helps individuals and organizations with decision making in their businesses.

Works on Social Network using Data Mining Techniques

Sentiment mining can be efficiently used to extract useful information from the huge volume of social network data. Several researches now concentrate on predicting the sentiment/opinion from images that are posted on social media. Quan Fang [2] proposed a model to predict opinions using the text and images posted on social networks like Flickr, trip advisor, etc.

1.2 Related Work in Online Product Recommendation

Zhang et al [9] proposed a method for ranking products using online reviews. In his work, subjective sentences and comparative sentences [from product comparison] are extracted from reviews, directed & a weighted product graph is constructed using dynamic programming and finally improved ranking algorithm is used to determine the rank of products.

Zhang et al [10] further proposed another feature based ranking product method. Firstly, all important features of a product are identified. Subjective sentences against each feature are evaluated and finally, overall ranking is provided based on each feature. Zhang et al [11] still improved the feature based ranking method by including influence factors using the number of helpful votes received for a review and review poster's age.

Yang Liu et al [12] proposed to rank products using online reviews by intuitionistic fuzzy set theory. The method uses two techniques: Sentiment analysis technique to identify the sentiment orientation for each product feature and ranking product technique which converts the sentiment orientation to fuzzy number and then an overall weightage is calculated for each alternative product using intuitionistic fuzzy weightage algorithm [IFWA]. Finally, ranking of alternative product is done using PROMETHE II.

Jinpeng Wang et al [13] proposed a new idea in the feature based ranking product method. In the proposed method, implicit demographic information is identified in acceptable volume from reviews and used to provide customer specific recommendation based on user category. Five major categories were identified based on age and sex of users – Children, Young female, Young male, Old male and Old female. Based on reviews, one more category Colleague was added. To identify the product adapter, unsupervised bootstrapping method was used based on the derived patterns from the large volume of data. Once the product adapter was categorized, user preference distribution and Category specific distribution were calculated. This provides the preference degree of the user over the candidate products. On comparing this approach with Matrix factorization which is a widely used method in product recommendation, this shows an improvement of 10% in recall and precision.

Suke Li *et al* [22] proposed a liner-based approach to rank the product feature based on customer reviews on the web. The introduced method was called as Double propagation based linear method. Approach contains two steps: Step 1) Double propagation algorithm - to extract the product feature and sentiment words. Step 2) Gradient descent algorithm - to calculate the weightage of the product feature and rank product based on overall product rating. Double propagation method is built on basis of high correlation of product feature and sentiment word. It uses extracted feature words to identify sentimental words and then from the extracted sentimental words, hidden product features are determined. It assumes that customers post reviews on a single product. Based on opinions, unit presence and negative indicator, weightage of the product feature is calculated. Overall predictive ranking of the product is obtained from overall product feature ranking present in the review.

Jayasekara and Wijayanayake [23] proposed the smiley based approach to generate opinion summary based on product features. The method extracts the product feature and its sentiment polarity. The method contains 3 steps: In Step1, all the product features are extracted. POS tagging is used to identify the frequently used product features. In Step 2, sentences with opinions and their polarity are identified in each review. Dictionary-based approach is used here. SentiWordNet is used to identify the sentiment polarity associated with the sentence. Smiley based approach is introduced for those sentences which were missed by the SentiWordNet approach. In Step3, the opinion results are summarized. The results show more accuracy than the conventional feature selection method.

III. COMPARATIVE ANALYSIS

Anuj Sharma and Shubhmaoy Dey [25] focused on two approaches - Machine learning feature selection and NPL Lexicon approach. Five Feature selection and three popular Sentiment lexicons are evaluated and classified with SVM. Results show that Gain Ratio performs well when a number of features are more and Information gain looks stable when compared to other techniques. Performance of classifier depends on the number of the prominent features selected. Lexicon result shows less performance when compared to machine language feature selection. **Table 2** shows the comparison of three sentiment lexicon.

Table 2: Sentiment Lexicon Comparison

Sentiment Lexicon	Recall	Precision	F-Measure
HM	0.49	0.64	0.55
GI	0.739	0.76	0.75
Opinion lexicon	0.675	0.723	0.70

Gautami et al [26] focused both Machine learning and NLP approach. Used higher order 4 gram for feature selection and compared with SVM and NB classification. Results show High order 4 gram feature better precision when compare to unigram and bi gram given in **Table 3**. SVM Classifier outperforms NB classification.

Table 3: N-gram comparison results

N gram	Accuracy	Precision	Recall
Unigram(n=1)	84.75	82.63	88.00
Bigram(n=2)	86.75	87.31	88.00
Trigram(n=3)	86.00	86.00	86.00
4 gram (n=4)	86.00	86.36	85.50

Rui Xia et al [14] proposed dual sentimental analysis method, which expands the data by creating a reversed shift review for each subjective training review. Dual training algorithm is proposed to consider reviews of both original and reverse side for learning a sentiment classifier. **Table 4** shows the classifier accuracy of DSA model improved with unigram and bigrams when compare to other methods.

Table 4: Dual Sentiment analysis (DSA) accuracy level comparison

Dataset	Baseline	LSS	DSA-WN
Book	0.779	0.809	0.823
DVD	0.801	0.823	0.831
Electronics	0.826	0.844	0.857

Cagatay Catal et al [15] proposed the benefits of using multiple classifier concepts on Turkish sentiment classification. The voting algorithm is used with three classifiers. The experimental results show the multiple classifier system outperforms single classifier technique. **Table 5** shows the accuracy of multiple classifier system with other single techniques.

Table 5: Multiple Classifier System accuracy comparison

Classifier	Accuracy (Book Reviews)	Movie Reviews Accuracy	Shopping Reviews Accuracy
Naive Bayes	85.48%	82.56	79.66
SVM	82.89	81.04	79.15
Multiple classifier System	86.13	83.68	79.96

Matthias Meiere at el [16] introduced leading and lagging information which helps to improve the model predictive performance. Information which is available even before the content is posted is called as leading information (e.g. user profile, previous post etc.). Lagging information is the information received after the content is posted (e.g. Comments, likes, retweet etc.). **Table 6** shows the increased accuracy when leading and lagging information are added.

Table 6: Accuracy level comparison when considering Leading and Lagging information.

Classifier / Accuracy level	User Post Variable	UPV +Leading	UPV +Leading + Lagging
Random Forest	0.771	0.796	0.835
SVM	0.724	0.762	0.804

The comparative analysis of product recommendation in the existing paper is given in **Table 7**. Proposed work involves using real-time social networking data (including other implicit information present in the reviews like occasions etc.) to identify customer specific recommendations.

To understand the different classification techniques, we have taken movie review dataset and performed different classification techniques. Observed the performance (accuracy level) of each classification is given in Table 8. It is elaborated more in Appendix A.

Table 8: Comparison of different classification for movie review dataset

S.No.	Learning Technique	Accuracy
1	Support Vector Machine	58.00%
2	Decision Tree	58.00%
3	Navie Bayes	50.00%
4	KNN	59.00%
5	Random Forest	54.00%

IV. CONCLUSION

Opinions and review comments for any product or service are available online and are treated as important assets for organizations to know more about consumer expectation. This information is also useful for consumers to make decisions before buying, by comparing and ranking alternative product features. It is practically impossible for anyone to read all reviews posted online manually. That's why the proposed system is of importance to compare review comments and come up with rankings for each alternative feature product, reducing consumers' overhead and simplifying their decision-making process.

V. Future Scope

The proposed approach can be further improved by including Fuzzy logic. By introducing Fuzzy logic, influencing factors like time and location, the accuracy of recommendation can be improved.

Appendix A. Comparison of different classification techniques

Considered movie review dataset and RapidMiner tool is used for comparing different classification techniques such as Naive Bayes, Super Vector Machine (SVM), Decision tree, Neural Network, KNN and Random Forest. Results are illustrated in the Table A. and observed KNN classification provides more accuracy when compare to other techniques. Before start the classification, preprocess steps: tokenize, remove stop words and stemming are applied to improve the performance.

Figure 1 : Preprocessing the input movie review dataset

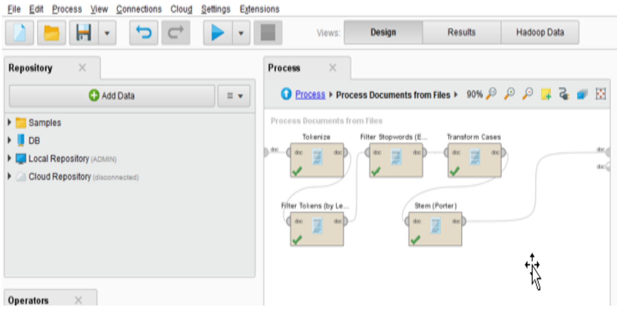

Figure 2 : Naive Bayes Classification

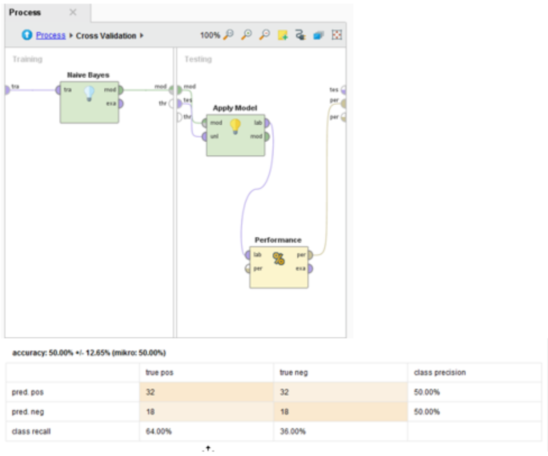

Figure 3 : SVM Classification

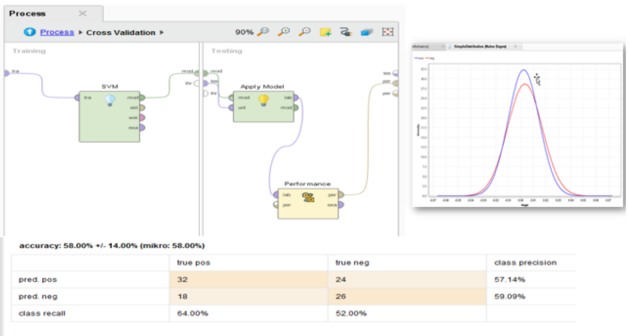

Figure 4 : Decision Tree Classification

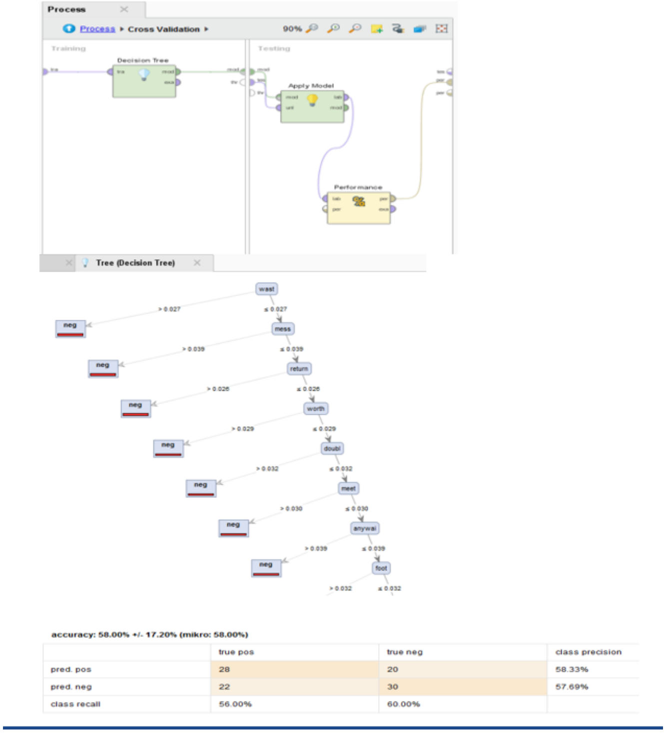

Figure 4 :KNN Classification

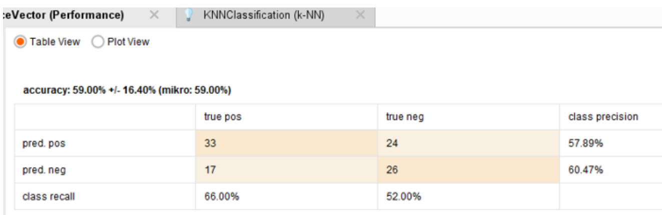

Figure 6 : Random Forest Classification

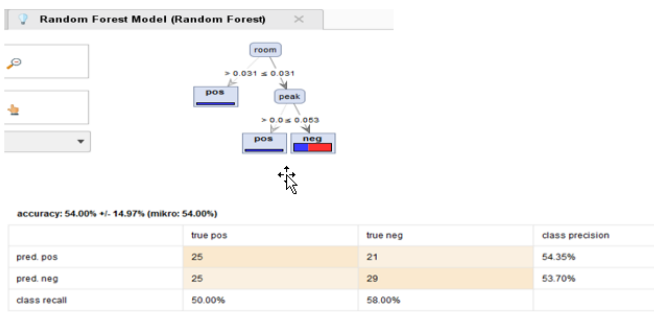

Figure 7 : Example dataset

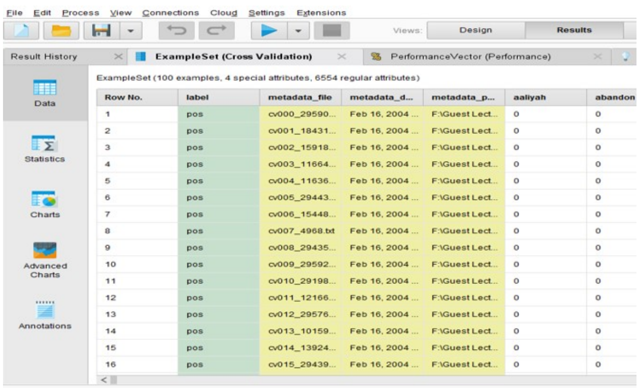

Table A.1. Comparison of accuracy level of different classification Techniques

S.No.	Learning Technique	Accuracy
1	Support Vector Machine	58.00%
2	Decision Tree	58.00%
3	Naive Bayes	50.00%
4	KNN	59.00%
5	Random Forest	54.00%

VI. REFERENCES

[1]. Francesco Corea,"Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector", Big Data Research, Vol. 4, pp. 70-74, 2016.

[2]. Quan Fang, Changsheng Xu, Jitao sang, Shamim Hossain M and Ghulam Muhummad, "Word of mouth understanding: entity centric multimodal aspect opining mining in Social media", IEEE Transactions on Multimedia, Vol.17, No.12, pp. 411-427, July 2016.

[3]. Xinjie Zhou, Xiaojun wan and Jianguo Xiao, "CMiner - Opinion extraction and summarization for Chinese microblogs", IEEE Transactions on Knowledge and Data Engineering", Vol. 28, pp. 7, July 2016.

[4]. Sanjiv R Das and Mike Y Chen, "Yahoo for Amazon - Sentimental extraction from small talk on the web", Management science, Vol.53, No.9, pp. 1375-1388, 2007.

[5]. B. Agarwal, N. Mittal, "Prominent Feature Extraction for Sentiment Analysis", Springer International Publishing, Socio-Affective Computing, Vol. 2, pp. 21-45, 2015.

[6]. Manizheh Ghaemi, Mohammad-Reza Feizi-Derakhshi, "Feature selection using Forest Optimization Algorithm", Pattern Recognition, Vol. 60, pp.121-129, 2016.

[7]. V Bolon Canedo and I Porto Diaz, "A Framework of cost based feature selection", Pattern Generation, Vol. 47, pp. 2481-2489, 2014.

[8]. Pawel Teisseyre, "Feature Ranking for multi label classification using Markov Networks", Neuro computing, Vol. 205, pp.439-454, 2016.

[9]. K. Zhang, R. Narayanan, A. Chaudhary, "Mining Online Customer Reviews For Ranking Products", EECS Department, Northwestern University, Technical Report, 2000.

[10]. K. Zhang, R. Narayanan, A. Chaudhary, "Voice of the customers: mining online customer reviews for product feature-based ranking", WOSN'10, Boston, USA, 2010.

[11]. K. Zhang, Y. Cheng, W.K. Liao, A. Chaudhary, "Mining millions of reviews: a technique to rank products based on importance of reviews", Proceedings of the 13th International Conference on Electronic Commerce, ACM, New York, USA, Article No.12, 2011.

[12]. Yang Liu, Jian -Wu and Zhi-Ping Fan, "Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory", Information Fusion, pp. 149-161

[13]. Jinpeng Wang, Wayne Xin Zhao, Yulan He and Xiaoming Li, "Leveraging product adapter information from Online reviews for product recommendation", AAAI Conference on web and social media, pp. 464-472, 2015.

[14]. Rui Xia, "Dual Sentiment Analysis", Information Fusion, 2015.

[15]. Cagatay Catal, Mehmet Nangir, "Sentiment classification model based on multiple classifier ", Applied Soft Computing, Vol. 50, pp. 135-141, November 2016.

[16]. Matthijis Meire, Michel Ballings and Dirik Van den Poel, "The added Value of auxiliary data in Sentiment analysis of Facebook posts", Decision Support Systems, Vol. 89, pp. 98-113, 2016.

[17]. Vimal Kumar B. Vaghela and Bhumika M. Jadav, "Analysis of Various Sentiment Classification Techniques", International journal of Computer applications, Vol. 140, pp. 22-27, April 2016.

[18]. Shehal A Mulay, Shrijeet J, Mohit R Shaha, Hrishikesh V Vibhute and Mahesh P, "Sentimental analysis and opinion mining with social networking - Predicting Box office collection of Movie", International journal of emerging research in management and technology, Vol.5, No.1, pp. 74-79, January 2016.

[19]. Sudipta Deb Roy and Sankar Kumar Chakraborty, "Impact of Social Media / Social Networks on Education and life", Vol. I, No. I, pp. 141-147, February 2015.

[20]. Thomas L Griffiths, Mark Steyvers, David M Blei and Joshua B Tenenbaum, "Integrating topics and syntax", In advance in neural information processing systems, MIT press, Vol. 17, pp. 537-544, 2005.

[21]. Shengsheng Qian and Tiazhu zing, "Multi modal event topic model for social event analysis", IEEE Transactions on Multimedia, Vol. 18, No. 2, pp. 233 - 246, Feb 2016.

[22]. Suke Li, Zhi Guan, Liyong Tang and Zhong Chen,"Exploiting Consumer Reviews for Product Feature Ranking", Journal of Computer Science and Technology, Vol. 27, No.3, pp. 635-649, July 2012.

[23]. I R Jayasekara and W M J I Wijayanayake, "Opinion mining of customer reviews: Feature and Simile based approach", International journal of Data mining & Knowledge Management process (IJKDP), Vol.6, No.1, pp.1-11, January 2016.

[24]. Olena Kummer,"Feature Selection in Sentiment Analysis", CORIA 2012, Bordeaux, Page No. 273-284

[25]. Anuj Sharma and Shubhmaoy Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis", Special issue of International Journal of Computer applications, pp. 15 - 20, June 2012.

[26]. Gautami Tripati and Naganna, "Feature Selection and Classification approach for Sentiment Analysis", International Journal of MILAIJ, Vol. 2, No. 2, pp. 1-12, June 2015.