

# Cross Lingual Opinion Mining using Machine Learning Techniques

<sup>[1]</sup> Abhiseha J, <sup>[2]</sup> Sahaana R T, <sup>[3]</sup> Dharshini K, <sup>[4]</sup> Menaha R  
Student of Information Technology MCET ,Pollachi Coimbatore ,India

**Abstract**— Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining is also called sentiment analysis that involves building a system to collect and categorize opinions about a product. Opinion mining can be used in several ways such as marketers in success of an ad campaign, new product launch, determine which versions of a product is popular, etc., The main idea of the project is to segregate the reviews as positive, negative and neutral and to give a clear idea about the product to the immigrants of France. The products are criticized and depicted as reviews in C discount, a French website. These reviews are categorized as positive, negative and neutral. But the main disadvantage is these reviews are only in French language. Thus reviews of other language are analyzed using machine learning techniques and translated to English for easy understanding of immigrants.

**Keywords**— Sentiment analysis, SWN Algorithm, Machine Learning

## I. INTRODUCTION

The online shopping websites are usually accomplished with the reviews. These reviews are categorized based on the ratings in some websites. In those websites if the reviews are classified then the users of the website can easily make choices on the product to be purchased. The website chosen in this project is C Discount. Here the reviews are not segregated as positive, negative and neutral. Only star rating is provided. The users will find difficult to purchase a product. Also the reviews are depicted only in French language and so the immigrant of France again makes purchasing more complex. Thus our idea is to implement machine learning techniques to translate the French reviews in English. The reviews of C Discount products are taken as dataset. The Pre-processing techniques are applied to those reviews and finally the chart is generated based on the features obtained. The remainder of this paper is structured as follows: Section 2 summarizes the related works on cross lingual opinion mining. Section 3 details the modelling process using SWN algorithm; and section 4 includes the experimental analysis in the process of separating the reviews as positive, negative and neutral. In section 5 the process in section 3 and section 4 were analysed for verification.

## II. RELATED WORK

According to Farman Ali, Kyung-Sup Kwaka and Yong-Gi Kimb[6] the raw reviews and sentences are maintained in a database. After applying the morphological and semantic analysis the extracted data can be further used

for the process of tokenization and word tagging. The features are extracted based on Fuzzy Domain Ontology(FDO). Then the feature review classification is done based on Support Vector Machine(SVM) and opinion mining is done based on FDO. Finally the fuzzy interface layer is computed and hotel polarity is calculated. Roshan Fernandes, and Rio D'Souza analysed[16] the twitter data for customer reviews based on different methods such as data collection, data pre-processing, data extraction, classification and evaluation. The pre-processing includes replacing emoticons, URL and hashtags, lowercase and identifying punctuations. The data extraction includes four steps such as POS tagging, tokenization, stop words and feature selection. After applying these methodologies the consolidated opinion about the various customer reviews are generated. Xiaowen Ding and Bing Liu[21] studied the problem of determining the semantic orientations (positive or negative) of opinions expressed on product features in reviews. Most existing approaches use a set of opinion words for the purpose. However, the semantic orientations of many words are context dependent .In this paper, they proposed to use some linguistic rules to deal with the problem together with a new opinion aggregation function. Extensive experiments show that these rules and the functions are highly effective. A system, called Opinion Observer, has also been built. Yai Wu and Martin Ester[23] revealed that aspect-based opinion mining from online reviews has attracted a lot of attention recently. Given a set of reviews, the main task of aspect-based opinion mining is to extract major aspects of the items and to infer the latent aspect ratings from each review. However, users may have different preferences

which might lead to different opinions on the same aspect of an item. Even if fine-grained aspect rating analysis is provided for each review, it is still difficult for a user to judge whether a specific aspect of an item meets his own expectation. In this paper, they studied the problem of estimating personalized sentiment polarities on different aspects of the items. They proposed a unified probabilistic model called Factorized Latent Aspect Model (FLAME), which combines the advantages of collaborative filtering and aspect based opinion mining. FLAME learns users' personalized preferences on different aspects from their past reviews, and predicts users' aspect ratings on new items by collective intelligence. Experiments on two online review datasets show that FLAME outperforms state-of-the-art methods on the tasks of aspect identification and aspect rating prediction. Diana Terezhnaya, Maruska Mascarenhas[5] proposed a novel approach to opinion mining in Konkani language. The proposed model which is called KOP in an opinion mining systems for texts written in Devnagiri scripted Konkani. It takes into consideration several important aspects in the language text such as negations, discourses, sarcasm and positional features, which influence the opinion being conveyed by the document text. It finally outputs the polarity of the document text as positive, negative or neutral. Anastasia Giachanou and Fabio Crestani[2] analyzed twitter in a two-fold process as it has become an important communication tool for people to share their opinions. The two-fold includes analysis of the topical distribution of tweets and opinion retrieval based on proximity. The majority of tweets include a single topic which include the short length of a tweet. Chun-Wen Li, Hui-Chi Chuang, and Sheng-Tun Li[4] make use of both opinion mining and hedonic analysis to reduce the demand of a commodity. They used smart phone market as a research target and proposed a framework to demonstrate their assumptions. This framework utilizes opinion mining techniques to extract the words and features. The managerial implications for firms and proper strategies are provided based on the results.

Geeta and Rajdeep Niyogi[7] found that great number of people use social media and going vocal with their thoughts. They collected tweets from twitter on major events using demographic analysis. By using this the users opinion, positive and negative sentiments differs. The sentiment analysis techniques are used for the analysis of the demographic behavior of users. Jagbir Kaur, and Meenakshi Bansal[8] used the trend of the

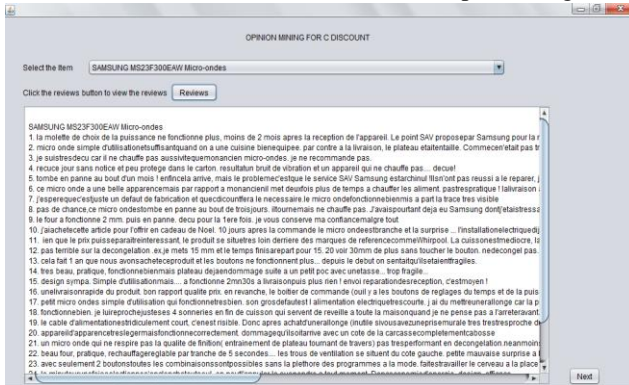
online hopping that has grown over the recent years. The product review report are generated based upon the multiple features altogether. The positive and negative points about each and every product play the significant role in the selection of products. Polarity based accuracy assessment is used for measuring the performance and accuracy instead of given parameters. KIM Yoosin, KIM Taeyun, PARK Miri, KANG Suna, CHOI Yiseul, CHOI Sanghyun[11] found the behavior of consumers from them to discover market intelligence and business insight using the modern opinion mining methods and techniques. Analyzing and comparing the consumers opinion helps to gain in-depth and fine-grained understanding on the human and social behavioral dynamics. The four phase methods such as aggregation, data qualification, analytics and presentation are used for the analysis of quantitative measurements. Nikos Tsirakis, Vasilis Pouloupoulos, Panagiotis Tsantilas, Iraklis Varlamis [13] studied about the PaloPro services. The background of PaloPro service includes aspect based opinion mining, competitive systems, PaloPro media analytic services, infrastructure and software to achieve the infrastructure. Scalability, System training and heterogeneity are some of the challenges in PaloPro process. Ratab Gulla, Umar Shoaiba, Saba Rasheedb, Washma Abidb, Beenish Zahoorb [15] use the techniques such as data extraction, data transformation, tweet cleaning, tweet parsing, database creation, collecting and organizing training data and supervise sentiment taggers on the political context of twitters data. Converting all tweets into lower case, removing emotions and punctuations, removing URL's are involved in tweet cleaning while extracting hash tags and extracting change of direction indicators are involved in tweet parsing. Shweta Rana and Archana Singh[19] consider the positive and negative sentiments using film user reviews for exploring the sentiments. Naïve Bayes, linear SVM and synthetic words are the algorithms used for performing the sentimental analysis. Data collection, mining and preprocessing are the three steps involved in sentimental analysis. The methodologies used are dataset, text preprocessing porter algorithm, linear support vector machine and naïve bayes. Linear SVM produces the best accuracy. In this paper, the reviews are collected from the French website C discount and the dataset for further processing is generated.

### III. PROPOSED SYSTEM

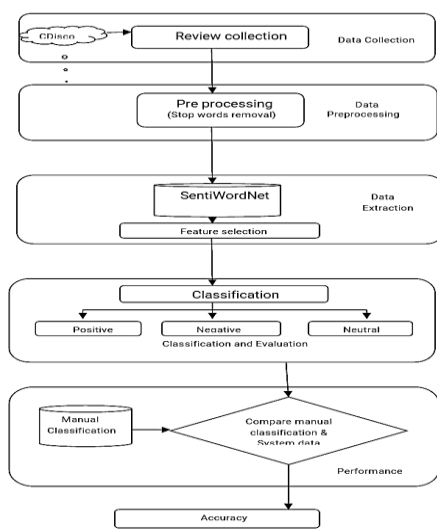
Collaborative filtering is used only to English language. In this system this technique is applied to other language such as French. The reviews are collected from the French website C discount and classified using SWN algorithm as positive, negative and neutral reviews. Thus this will be helpful for immigrant of France to use this website.

#### A. Review Extraction

The reviews are collected from a French online shopping website called C Discount. Here the reviews are given rating but not segregated as positive, negative and neutral. Also the number of reviews for each product varies, as a consideration about 25 reviews from each product are extracted. Thus the dataset is collected for the products in C Discount and the reviews are extracted for processing.



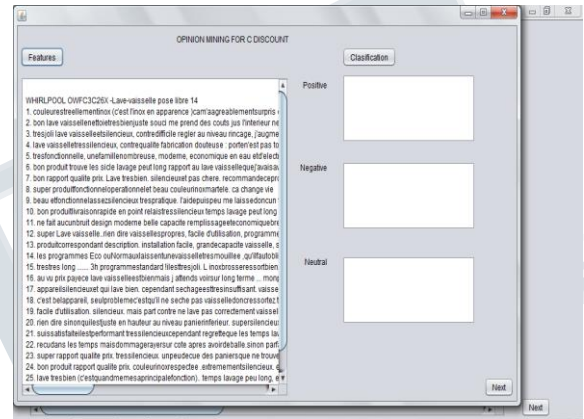
**Fig. 1. Review Collection for C Discount**



**Fig. 2. Proposed System Architecture**

#### B. Feature Extraction

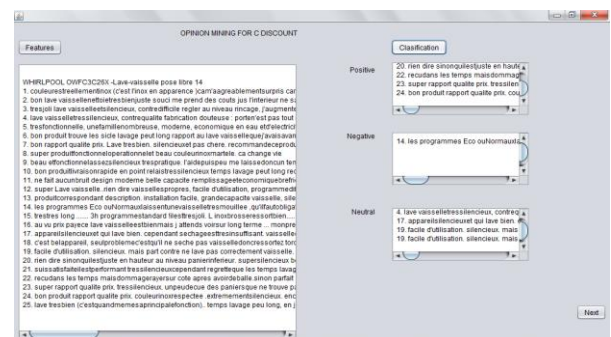
The feature for each product is extracted by removing the stop words from each review. The features are the emphasized words that can be taken as a clue for classification as positive, negative or a neutral word. The stop words are non-emphasized words that do not change the meaning of the review even when removed. These stop words are removed from the review set and the features of review are taken for processing. Thus from the extraction of feature words the review can be easily classified.



**Fig.3. Feature Extraction for a product**

#### C. Classification

The classification is done based on SWN algorithm where the reviews are separated as lines and the again separated as individual words for comparing the word to be positive, negative or neutral. These reviews are also compared with manual classification and verified for accuracy. Thus the classification is done and the reviews are separated as positive, negative and neutral.



**Fig. 4. Classification of Reviews**



#### IV. EXPERIMENTAL ANALYSIS

The product review analysis using the opinion mining approach plays the vital role in the classification and rating of the products. The French website C Discount is selected for the process of review collection. In this website the product reviews are classified only based on the star ratings. It will not be clear for the product selection. But in other shopping websites like Flipkart, Amazon, ebay the product reviews are classified as positive, negative and neutral. First, twenty electronic products are selected. For each product 40 to 50 reviews are collected. Then a JFrame form is designed for easy processing and analysis of the reviews. The collected reviews are maintained as a dataset. Each product review is stored in different document. Those reviews are called in the java program by using the specified path of each product. In the JFrame form these reviews are displayed when the review button is clicked. Secondly, the features are extracted for each and every product. For feature extraction the bag of words method is used. The set of stop words are collected and those stop words are removed from each and every review. The SentiWordNet(SWN) algorithm is used for comparing the stop words and reviews. In this SWN algorithm the total file is being segregated as paragraphs. Next the paragraph is segregates as lines and then the lines are segregated as words. Then those words are compared with the collected stop words. When the feature button is selected the stop words are removed and the features are displayed for the specified product. The feature extraction and classification of the reviews are done in another JFrame. The positive and negative words in French are collected manually. These words are stored in separate arrays. Again the bag of words method and SWN algorithm is used and the reviews are classified as positive, negative and neutral respectively. When the classification button is selected the reviews are classified and displayed in the respective textboxes for the specified product. Simultaneously the reviews are classified as positive, negative and neutral for each product manually. The manual classification of the reviews and the experimental classification are compared and the accuracy is found using it. Next the French review classification is translated as English review classification using bag of words method and SWN algorithm. For this translation a new JFrame form is being designed. When the translate button is clicked, the translated reviews are classified and

displayed. The comparison for positive, negative and neutral reviews is displayed clearly using a jFree chart. When the chart button is clicked, the chart is generated and displayed for a specific product.

S.NO	Product	Positive	Negative	Neutral
1.	LOFT Table a manger de 8 a 10 personnes style indu	8. Impeccable ne suis pas decu top contre aucunes planches n'est numerote montage on pas trop 2 faire	4. table arrivee les temps seul probleme c'est que trous mal pre perces (obligé ne percer meme) les trous pre pperces decalles obligé reajuster tout ! peu bricoleuse j'ai fini abimer table ! Une vraie galere. !!! Tres decue service !	1. table ne ressemble pas photo. plateau gris non pas couleur chene naturel. photo laisse penser plateau epais mais pas tout. Bref grosse deception, demande retour en cours.
16.	Table design necessite espace car grande Pas regret cet achat rapport qualite prix indiscutable	21. Tres pratique repas en famille (grande famille!!) Pour 8 personnes c'est largement suffisant. Je conseille fortement.	6. Il n'y pas notice montage produit trop lourd difficile monte	3. Les pieds ne pas en metal mais en imitation bois pas ressemblant photo difficile det tenir 10
23.	able grande on peut mettre 10 personnes manger sans serrer dessus manger ne regrette pas l'avis acheter c discount	25. Rendu sympa. montage facile. Bien haute que ma precedente table coup les chaises que venais d'acheter ne conviennent plus		5. Je suis plutot decus article parcequ'il n'est écrit nule part que c'est plateau fin des entourage monter non gros plateau on voit les photo pareil les pieds , donc pas content
				7. Joli table!! sinon dommage l'angle ma table abime.
				9. Belle table mais defaut plateau reception; table n'a pas renvoyee car cela faisait deja d'un mois que commande avait faite l'attente longue.!
				10. Je voyais moins grande pieds brillants mais non! Ca reste jolie table... Facile monter solide...

**Table 1. Manual classification of reviews for a product**

#### V. RESULTS AND DISCUSSIONS

Thus the review collection on various products, feature extraction and classification of reviews such as positive, negative and neutral and translating them to English language has been performed and their accuracy has been verified using manual classification.

#### VI. CONCLUSION

The proposed scenario can be viewed as a necessary implementation for the immigrants of France. Thus the people who are immigrants to France can well utilize the website for on-line purchasing. It is useful in the way it is less time consuming as the best products are easily visualized as positive and negative. Thus on-line purchasing is made convenient by the way it depicts the reviews in a simple manner.

#### REFERENCES

- [1] AliakseiSeveryn, Alessandro Moschitti, Olga Uryupina, Barbara Plank, KatjaFilippova, "Multi-lingual opinion mining on YouTube", 0306-4573/ 2015 Elsevier Ltd.
- [2] Anastasia Giachanou and Fabio Crestani, "Opinion Retrieval in Twitter: Is Proximity Effective?", ACM. ISBN 978-1-4503-3739:1146-1151,2016.

- 
- [3] A.Suriyal and Dr. M. Prabakaran, "Opinion mining and sentiment analysis using data mining techniques-A Survey".
- [4] Chun-Wen Li, Hui-Chi Chuang, and Sheng-Tun Li, "Hedonic Analysis for Consumer Electronics Using Online Product Reviews", Proceedings of the IEEE 5th IIAI International Congress on Advanced Applied Informatics, 2016.
- [5] Diana Terezinha Miranda, Maruska Mascarenhas, "KOP: An Opinion Mining System in Konkani", Proceedings of IEEE International Conference On Recent Trends In Electronics Information Communication Technology, pages 702-206, May 20-21, 2016, India.
- [6] Farman Ali, Kyung-Sup Kwaka, Yong-Gi Kimb, "Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification", 1568-4946/ 2016 Elsevier B.V.
- [7] Geeta and Rajdeep Niyogi, "Demographic Analysis of Twitter Users", Indian Institute of Technology Roorkee, Proceeding of 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.
- [8] Jagbir Kaur, and Meenakshi Bansal, "Multi-Layered Sentiment Analytical Model for Product Review Mining", Proceedings of Fourth International Conference on Parallel, Distributed and Grid Computing, pages 413-420, YCOE, Talwandi Sabo, Punjab, India, 2016.
- [9] Jumayel Islam, Zubair Azami Badhon and Pintu Chandra Shill, "An Effective Approach of Intrinsic and Extrinsic Domain Relevance Technique for Feature Extraction in Opinion Mining", Khulna University of Engineering & Technology, Proceedings of 5th International Conference on Informatics, Electronics and Vision (ICIEV), pages 428-432, 2016.
- [10] Jyoti S. Deshmukh, Amiya Kumar Tripathy, "Entropy based classifier for Cross-Domain opinion mining", S2210-8327(17)30096-9.
- [11] KIM Yoosin, KIM Taeyun, PARK Miri, KANG Suna, CHOI Yiseul, CHOI Sanghyun, "Sentiment Analysis of Consumer Opinion in Blogs: A Case Study in Ramen Market, Jeju Island", Republic of Korea © 2015 ACM. ISBN 978-1-4503-3846, 2015.
- [12] Necmiye Genc-Nayebi, Alain Abran, "A Systematic Literature Review: Opinion Mining Studies from Mobile App Store User Reviews", S0164-1212(16)30229-1.
- [13] Nikos Tsirakis, Vasilis Pouloupoulos, Panagiotis Tsantilas, Iraklis Varlamis, "Large scale opinion mining for social, news and blog data", 0164-1212/ 2016 Elsevier Inc.
- [14] Pankaj Gupta, Ritu Tiwari and Nirmal Robert, "Sentiment Analysis and Text Summarization of Online Reviews: A Survey", Proceedings of International Conference on Communication and Signal Processing, pages 0241-0247, April 6-8, 2016, India.
- [15] Ratab Gulla, Umar Shoaiba, Saba Rasheedb, Washma Abidb, Beenish Zahoorb, "Pre Processing of Twitter's Data for Opinion Mining in Political Context", 1877-0509/ 2016 Elsevier B.V.
- [16] Roshan Fernandes, and Rio D'Souza, "Analysis of Product Twitter Data through Opinion Mining", St. Joseph Engineering College Vamanjoor, India, 2016.
- [17] R. Piryani, D. Madhavi, V.K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000-2015", 0306-4573/ 2016 Elsevier Ltd.
- [18] Shiliang Sun, Chen Luo, Junyu Chen, "A Review of Natural Language Processing Techniques for Opinion Mining Systems", S1566-2535(16)30111-7.
- [19] Shweta Rana and Archana Singh, "Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques", ASET-IT, Amity University Uttar Pradesh, Proceedings of 2nd International Conference on Next Generation Computing Technologies (NGCT-2016), pages 106-112, Dehradun, 2016.
- [20] Walid Cherif, Abdellah Madani, and Mohamed Kissi, "A combination of Low-level light stemming and
-

Support Vector Machines for the classification of Arabic opinions”, University Chouaib Doukkali, El Jadida, Morocco, 2016.

[21] Xiaowen Ding and Bing Liu, “The Utility of Linguistic Rules in Opinion Mining”, ACM 978-1-59593-597:811-812,2015.

[22] Xinjie Zhou, Xiaojun Wan\* and Jianguo Xiao, “CLOpinionMiner: Opinion Target Extraction in a Cross-Language Scenario”.

[23] Yai Wu and Martin Ester, “FLAME: A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering”, ACM-2015.

[24] Zheng Yan, Xu-yang Jing, WitoldPedrycz, “Fusing and Mining Opinions for Reputation Generation”, S1566-2535(16)30159-2.

